

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2002, 13:03:56 ; Search time 1551.88 Seconds
(without alignments)
2041.046 Million cell updates/sec

Title: US-09-816-391A-1_COPY_1_192

Sequence: 1 gctggggcggcggcgcacatga.....tgaccacagaagatgcttt 192

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

GenBank: *
1: gb_ba: *
2: gb_hng: *
3: gb_in: *
4: gb_com: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_om: *
20: em_or: *
21: em_ov: *
22: em_pat: *
23: em_ph: *
24: em_pl: *
25: em_ro: *
26: em_sts: *
27: em_sy: *
28: em_un: *
29: em_vl: *
30: em_hng_hum: *
31: em_hng_in: *
32: em_hng_rod: *
33: em_hng_hum: *
34: em_hng_in: *
35: em_hng_rod: *
36: em_hng_other: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35.8	18.6	171413	2	AC092873 Pan trogl
2	34.8	18.1	152423	9	AC087432 Homo sapi
3	34.8	18.1	160410	2	AC020753 Homo sapi
4	34.8	18.1	166156	9	AC018814 Homo sapi
5	34.8	18.1	166156	17	AC087085 Homo sapi
6	34.8	18.1	181870	2	AC024216 Homo sapi
7	34.8	18.1	188432	2	AC018497 Homo sapi
8	34.4	17.9	148269	2	AC026486 Homo sapi
9	34.4	17.9	162249	2	AC020682 Homo sapi
10	34.4	17.9	166973	9	AL451049 Human DNA
11	34.4	17.9	188026	2	AL356952 Homo sapi
12	34.2	17.8	155711	3	VIYVD10
13	34.2	17.7	57861	2	AC023713 Homo sapi
14	34	17.7	166872	9	AC008554 Homo sapi
15	34	17.7	179393	9	AC010636 Homo sapi
16	33.6	17.5	17073	3	AC084160 Caenorhab
17	33.4	17.4	14784	9	HS1189824 Human DNA
18	33.4	17.4	192711	2	AL590139 Homo sapi
19	33.2	17.3	630	11	G54937 SHGC-100131
20	33.2	17.3	74436	2	AC026625 Homo sapi
21	33.2	17.3	87258	9	AF091512 Homo sapi
22	33.2	17.3	132910	2	AC014319 Drosophila
23	33.2	17.3	151097	2	AC083846 Homo sapi
24	33.2	17.3	168700	2	AC009115 Homo sapi
25	33.2	17.3	181021	9	AP001781 Homo sapi
26	33.2	17.3	182776	2	AC009766 Homo sapi
27	33.2	17.3	212370	3	DMBR744
28	33.2	17.3	239566	9	HSJ03147 Homo sapi
29	33	17.2	207588	2	AC021632.1 Continuation (2 of
30	33	17.2	207588	2	AC087162 Mus muscu
31	33	17.2	264893	2	AC019028 Mus muscu
32	32.8	17.1	35284	2	CEH27E01 Mus muscu
33	32.8	17.1	147986	2	AC083836 Homo sapi
34	32.8	17.1	167205	2	AC023871 Homo sapi
35	32.8	17.1	185252	2	AC026561 Homo sapi
36	32.6	17.0	127206	3	CEY53F48 Caenorhab
37	32.6	17.0	157534	2	AC090100 Homo sapi
38	32.6	17.0	170820	2	AC009855 Homo sapi
39	32.6	17.0	175697	2	AP002512 Homo sapi
40	32.6	17.0	256172	2	AC005139 Plasmodiu
41	32.4	16.9	375	5	MFT12RNA X82513 F. tropica m
42	32.4	16.9	27281	2	AC020340 Drosophila
43	32.4	16.9	108924	3	AC005889 Drosophila
44	32.4	16.9	134963	9	HS119916 Human DNA
45	32.4	16.9	164179	2	AL445533 Homo sapi

ALIGNMENTS

RESULT 1
AC092873 171413 bp DNA HTG 01-AUG-2001
LOCUS AC092873/c
DEFINITION Pan troglodytes clone RP43-22112, WORKING DRAFT SEQUENCE, 5
unordered pieces.
ACCESSION AC092873
VERSION AC092873.1 GI:15055306
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE chimpanzee.
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.
REFERENCE 1 (bases 1 to 171413)
Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,
Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S., Dietrich,N.L.,
Granite,S., Guan,X., Gupta,J., Ho,S.-L., Idol,J.R., Karlins,E.,
Lee-Il-in,S.-O., Legaspi,R., Lim,M., Maduro,Q.L., Maduro,V.B.,
Masello,C., Mastrian,J.D., McCloskey,J.C., McDowell,J.,
Pearson,R., Prasad,A., Shevchenko,Y., Snyder,B., Stantlipp,S.,
Thomas,J.W., Thomas,P.J., Touchman,J.W., Tsurgon,C., Vogt,J.L.,

```

TITLE      Walker,M.A., Metheby,K.D., Zhang,L.-H. and Green,E.D.
JOURNAL    NISC Comparative Sequencing Initiative
REFERENCE  Unpublished
AUTHORS    2 (bases 1 to 171413)
           Green,E.D.
JOURNAL    Direct Submission
           Submitted (01-AUG-2001) NIH Intramural Sequencing Center, 8717
           Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT    ----- Genome Center
           Center: NIH Intramural Sequencing Center
           Center code: NISC
           Web site: http://www.nisc.nih.gov
           Contact: nisc.mouse@nhgri.nih.gov
           ----- Project Information
           Center project name: ani
           Center clone name: 022112
           ----- Summary Statistics
           Sequencing vector: plasmid; n/a; 100% of reads
           Chemistry: Dye-terminator Big Dye; 100% of reads
           Assembly program: Phrap; version 0.990319
           Consensus quality: 170160 bases at least Q40
           Consensus quality: 170445 bases at least Q30
           Consensus quality: 170654 bases at least Q20
           Insert size: 152000; agarose-fp
           Insert size: 168000; pulse-field-gel
           Insert size: 171013; sum-of-contigs
           Quality coverage: 15.61x in Q20 bases; agarose-fp
           Quality coverage: 14.12x in Q20 bases; pulse-field-gel
           Quality coverage: 13.87x in Q20 bases; sum-of-contigs
           -----
           * NOTE: This is a 'working draft' sequence. It currently
           * consists of 5 contigs. The true order of the pieces
           * is not known and their order in this sequence record is
           * arbitrary. Gaps between the contigs are represented as
           * runs of N, but the exact sizes of the gaps are unknown.
           * This record will be updated with the finished sequence
           * as soon as it is available and the accession number will
           * be preserved.
           *
           *   1       2901: contig of 2901 bp in length
           *   2902     3001: gap of unknown length
           *   3002     7859: contig of 4858 bp in length
           *   7860     7959: gap of unknown length
           *   7960     17459: contig of 9500 bp in length
           *   17460    17559: gap of unknown length
           *   17560    90578: contig of 73019 bp in length
           *   90579    90678: gap of unknown length
           *   90679    171413: contig of 80735 bp in length.
           *
           Location/unallifiers
           1. 171413
           /organism="Pan troglodytes"
           /db_xref="taxon:9598"
           /clone="RP43-22112"
           /clone_1ID="RP43"
           1. 2901
           /note="assembly-fragment"
           clone_end:T7
           vector_side:left"
           3002. .7859
           /note="assembly-fragment"
           7960. .17459
           /note="assembly-fragment"
           clone_end:SP6
           vector_side:right"
           17560. .90578
           /note="assembly-fragment"
           90679. .171413
           /note="assembly-fragment"
           BASE COUNT  53796 a 33490 g 51618 t          406 others
ORIGIN
Query Match      18.66; Score 35.8; DB 2; Length 171413;
Best Local Similarity 63.2%; Pred. No. 1.8;

```

Matches	55:	Conservative	0:	Mismatches	32:	Indels	0:	Gaps	0:
Oy	15	cctaagaatgccttgacacaagcataaatcctgctcgtatccgatattttaaacttcggg	74						
Db	113527	CCAGGCACTGGGTACATNAGTATCATATTTTTTTAAATTTCCTAAGTTTTAAAAACAATTTCAT	113468						
Oy	75	gaataagatgtcaaacacctataaaa	101						
Db	113467	ATTAAGCATTTTTGAAAACACTTAAAAA	113441						

RESULT	2		PRI.	11-FEB-2001
LOCUS	AC087432			
DEFINITION	Homo sapiens chromosome 3 clone RP11-96G4 map 3p, complete sequence.			
ACCESSION	AC087432			
VERSION	AC087432.2			
KEYWORDS	HTG.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
AUTHORS	1 (bases 1 to 152423) Xiong,H., Zhou,Y., Dong,H., Lin,W., Chen,B., Zhang,C., Zhang,Y., Cai,Z., Ying,H.F., Wang,H., Gu,W., Zhu,G., Tu,Y., Zhang,X., Jia,J., Shen,H., Zhang,D., Wu,C., Lu,G., Zhong,M., Jiang,H., Ren,S., Fu,G., Chen,Z., and Huang,M.			
TITLE	Chromosome 3p genomic sequence			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 152423)			
AUTHORS	Liu,B., Bao,J., Bao,Q., Bao,W., Bian,X., Cao,T., Chen,C., Chen,J., Ding,H., Dong,W., Fan,H., Feng,X., Gong,J., Guan,Q., Gu,X., Guo,D., He,L., Hu,S., Huang,F., Jin,Y., Kang,N., Li,C., Li,C., Li,G., Li,J., Li,L., Li,S., Li,T., Liu,Y., Liu,N., Liu,Y., Li,W., Li,W., Luo,D., Niu,Y., Qi,Q., Qi,x, Song,L., Song,S., Sun,M., Sun,W., Tan,X., Tao,R., Wang,H., Wang,J., Wang,L., Wang,L., Wang,R., Wang,X., Wang,X., Yao,B., Zeng,Y., Zhang,G., Xuan,Z., Xue,Y., Yan,C., Yang,X., Yu,B., Zeng,Y., Zhang,X., Zhang,H., Zhang,H., Zhang,L., Zhang,M., Zhang,X., Zhang,Y., Zhang,Y., Zhang,Z., Zhu,B., Zhu,N., Yu,J., and Yang,H.			
TITLE	Direct Submission			
JOURNAL	Submitted (05-JAN-2001) 1. Chinese Human Genome Center at Shanghai			
REFERENCE	2. Institute of Genetics, Chinese Academy of Sciences, P.R.China			
AUTHORS	3 (bases 1 to 152423) Bao,J., Bao,Q., Bao,W., Bian,X., Cao,T., Chen,C., Chen,J., Ding,H., Dong,W., Fan,H., Feng,X., Gong,J., Guan,Q., Gu,X., Guo,D., Guo,Z., He,L., Hu,S., Huang,F., Jin,Y., Kang,N., Li,C., Li,C., Li,F., Li,G., Li,J., Li,L., Li,S., Li,T., Liu,Y., Liu,N., Liu,B., Liu,Y., Li,W., Li,W., Li,Y., Luo,C., Luo,J., Luo,Y., Qi,Q., Qi.x, Song,L., Song,S., Sun,M., Sun,W., Sun,Y., Tan,X., Tao,R., Wang,H., Wang,J., Wang,J., Wang,J., Wang,L., Wang,L., Wang,R., Wang,X., Wang,X., Wang,Y., Wu,D., Wu.Q., Xie,F., Xuan,Z., Xue,Y., Yan.C., Yang.X., Yu.B., Zeng.Y., Zhang.G., Zhang.H., Zhang.H., Zhang.L., Zhang.M., Zhang.X., Zhang,X., Zhang,Y., Zhang,Y., Zhang,Z., Zhu.B., Zhu.N., Yu.J., and Yang,H.			

TITLE	JOURNAL
COMMENT	On Feb 11, 2001 this sequence version replaced gi:12039244.

TITLE	JOURNAL
COMMENT	Genome Center Center:Beijing Center Center code:Beijing Website:http://hgsc.igtp.ac.cn http://www.genomics.org.cn Contact:ngsc@igtp.ac.cn ----- Project Information Center project name:1% project Center clone name: RP11-96G4 ----- Summary Statistics Sequencing vector: pUC18; 100% of reads Chemistry: Dye-terminator; ET 5% of reads

on Feb 23, 2000 this sequence version replaced by: 06684192

* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

REFERENCE AUTHORS	TITLE
1 (bases 1 to 166156)	Chromosome 3p genomic sequence
Xiong,H., Zhou,Y., Dong,H., Lin,W., Chen,B., Zhang,C., Zhang,Y., Cai,Z., Ying,H.F., Wang,H., Gu,W., Zhu,G., Tu,Y., Zhang,X., Jia,J., Shen,H., Zhang,D., Wu,C., Lu,G., Zhong,M., Jiang,H., Ren,S., Fu,G., Chen,Z., and Huang,M.	
2 (bases 1 to 166156)	Unpublished
Li,G., Hu,S., Dong,M., Wang,J., Zhang,Y., Zhang,H., Liu,B., Bao,M., Sun,Y., Wu,Q., Wang,H., Yang,X., Cheng,C., Wang,Y., Niu,Y., Qi,X., Li,T., Zhang,H., Liu,N., Wu,D., Yu,B., Fan,H., Liu,Y., Wang,X.,	

TITLE	Direct Submission
JOURNAL	Submitted (21-DEC-1999) Human Genomic Center, Institute of Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing 100101, P.R.China
REFERENCE	3 (bases 1 to 166156)
AUTHORS	Zhou, Y., Xiong, H., Dong, H., Lin, M., Chen, B., Zhang, C., Zhang, Y., Cai, Z., Ying, H. F., Wang, H., Gu, W., Zhu, G., Tu, Y., Zhang, X., Jia, J., Shen, H., Zhang, D., Wu, C., Lu, G., Zhong, M., Jiang, H., Ren, S., Fu, G., Chen, Z., and Huang, M.
TITLE	Direct Submission
JOURNAL	Submitted (27-DEC-2000) 1. Chinese Human Genome Center at Shanghai 2. Institute of Genetics, Chinese Academy of Sciences, P.R.China 4 (bases 1 to 166156)
REFERENCE	Bao, J., Bao, Q., Bao, W., Bian, X., Cao, T., Chen, C., Chen, J., Ding, H., Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q., Gu, X., Guo, D., Guo, Z., He, L., Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, C., Li, F., Li, G., Li, J., Li, L., Li, S., Li, T., Liu, Y., Liu, N., Liu, B., Liu, Y., Li, W., Li, W., Li, Y., Luo, C., Luo, J., Niu, Y., Qi, Q., Qi, X., Song, L., Song, S., Sun, M., Sun, W., Sun, Y., Tan, X., Tao, R., Wang, H., Wang, J., Wang, J., Wang, J., Wang, L., Wang, R., Wang, X., Wang, X., Wang, Y., Wu, D., Wu, Q., Xie, F., Xian, Z., Xue, Y., Yan, C., Yang, X., Yu, B., Zeng, Y., Zhang, G., Zhang, H., Zhang, H., Zhang, L., Zhang, W., Zhang, X., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Zhu, N., Yu, J., and Yang, H.
TITLE	Direct Submission
JOURNAL	Submitted (06-MAR-2001) Human Genomic Center, Institute of Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing 100101, P.R.China
COMMENT	On Mar 6, 2001 this sequence version replaced gi:11992941.
TITLE	Genome Center
JOURNAL	Center:Beijing Center Center code:Beijing Website:http://hgsc.igtp.ac.cn http://www.genomics.org.cn Contact:ngc@igtp.ac.cn
TITLE	Project Information
JOURNAL	Center project name:1% project Center clone name: RP11-204C23
TITLE	Summary Statistics
JOURNAL	Sequencing vector: pUC18; 100% of reads Chemistry: Dye-terminator; ET 5% of reads Chemistry: Dye-terminator Big Dye; 4% of reads Assembly program: Phrap; version 0.990329 Consensus quality: 314 bases at least Q40 Consensus quality: 468 bases at least Q30 Consensus quality: 518 bases at least Q20 Insert size: 534; sum-of-contigs Quality coverage: 4.75x in Q20 bases;sum-of-contigs
FEATURES	Location/Qualifiers
SOURCE	1. 166156 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="3" /map="3p" /clone="RP11-204C23"
BASE COUNT	55129 a 31315 c 30797 g 48915 t
ORIGIN	
Query Match	18.1%; Score 34.8; DB 9; Length 166156;
Best Local Similarity	65.4%; Pred. No. 3.8;
Matches 51; Conservativity 0; Mismatches 27; Indels 0; Gaps 0;	
47	tgattcgtcttttcaataaccttcggggaatagatgtgaaacctataaacgcg 106
DB 108901	TGAAAGACGCTATTTTGTACACCTTGTGCACATATATTTTGAAGAACTACATGAAATACAA 108960
107	gtttcgcagaacatgc 124

Db	108961	TATTCCTAGAAAAATAC	108978
RESULT	5		
AC087085			
ID	AC087085	standard; DNA; HDU; 166156 BP.	
XX			
AC	AC087085;		
SV	AC087085.1		
XX			
DT	07-DEC-2000 (Rel. 66, Created)		
DT	07-DEC-2000 (Rel. 66, Last updated, Version 1)		
XX			
DE	Homo sapiens chromosome 3p clone RP11-204c23, complete sequence.		
XX			
XX	HTG.		
XX			
OC	Homo sapiens (human)		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia		
OC	Eutheria; Primates; Catarrhini; Hominoidea; Homo.		
XX			
RP	[1]		
RA	1-166156		
RA	Zhou Y., Xiong H., Dong H., Lin W., Chen B., Zhang C., Zhang Y., Cai Z.,		
RA	Yang D., Wu C., Lu G., Zhong M., Jiang H., Ren S., Fu G., Chen Z.,		
RA	Ying H.F., Wang H., Gu W., Zhu G., Tu Y., Zhang X., Jia J., Shen H.,		
RA	Zhang D., Wu C., Lu G., Zhong M., Jiang H., Ren S., Fu G., Chen Z.,		
RA	Huang M.;		
RT	"Chromosome 3p genomic sequence";		
RL	unpublished.		
XX			
RN	[2]		
RP	1-166156		
RA	Zhou Y., Xiong H., Dong H., Lin W., Chen B., Zhang C., Zhang Y., Cai Z.,		
RA	Yang D., Wu C., Lu G., Zhong M., Jiang H., Ren S., Fu G., Chen Z.,		
RA	Ying H.F., Wang H., Gu W., Zhu G., Tu Y., Zhang X., Jia J., Shen H.,		
RA	Zhang D., Wu C., Lu G., Zhong M., Jiang H., Ren S., Fu G., Chen Z.,		
RA	Huang M.;		
RT	Submitted (06-DEC-2000) to the EMBL/GenBank/DBJ databases.		
RL	1. Chinese Human Genome Center at Shanghai 2. Institute of Genetics,		
RL	Chinese Academy of Sciences, P.R.China		
XX			
CC	-----Genome Center-----		
CC	Center:Beijing Center Center code:Beijing		
CC	Website:http://www.hgc.sh.cn http://hgc.igtp.ac.cn		
CC	http://www.genomics.org.cn Contact:hgc@igtp.		
CC	NOTE: This record was originally submitted by Chinese Human Genome		
CC	Center at Shanghai (CHGC), and now maintained by CHGC, submitted by		
CC	Beijing Center.		
XX			
XX			
FH	Key	Location/Qualifiers	
FT	source	1. 166156	
FT		/chromosome="3p"	
FT		/db_xref="taxon:9606"	
FT		/organism="Homo sapiens"	
FT		/clone="RP11-82c3"	
FT		/clone="RP11-204c23"	
FT		1030. .1056	
FT	repeat_region	/rpt_family="Simple-repeat"	
FT		/rpt_unit="(GAAA)n	
FT		1250. .1371	
FT	repeat_region	/rpt_family="SINE/MIR"	
FT		/rpt_unit=MIR	
FT		1647. .1989	
FT	repeat_region	/rpt_family="LINE/L1"	
FT		/rpt_unit=LINE4	
FT		2019. .2158	
FT	repeat_region	/rpt_family="LTR/Retroviral"	
FT		/rpt_unit=LTRFB	
FT		2167. .2705	
FT	repeat_region	/rpt_family="LINE/L1"	
FT		/rpt_unit=LINEC3	

```
FT      repeat_region      2709..3765      /rpt_family="LINE/L1"
FT      /rpt_unit-L1MC3
FT      repeat_region      3766..3803      /rpt_unit-L1MC3
FT      /rpt_family="Simple_repeat"
FT      /rpt_unit-(CA)n
FT      repeat_region      3804..4627      /rpt_family="LINE/L1"
FT      /rpt_unit-L1MC3
FT      repeat_region      4664..4791      /rpt_family="LINE/L1"
FT      /rpt_unit-L1MC3
FT      repeat_region      4819..4980      /rpt_family="SINE/MIR"
FT      /rpt_unit-MIR
FT      repeat_region      complement(5935..6051)
FT      /rpt_family="LINE/L2"
FT      /rpt_unit-L2
FT      repeat_region      complement(6792..9493)
FT      /rpt_family="LINE/L1"
FT      /rpt_unit-L1PA5
FT      repeat_region      complement(9544..9591)
FT      /rpt_family="DNA/Mariner"
FT      /rpt_unit-MADE1
FT      repeat_region      10733..10885      /rpt_family="DNA/MER1_type"
FT      /rpt_unit-MER63A
FT      repeat_region      complement(13506..13551)
FT      /rpt_family="SINE/MIR"
FT      /rpt_unit-MIR
FT      repeat_region      14554..14587      /rpt_family="Low_complexity"
FT      /rpt_unit-AT-rich
FT      repeat_region      complement(14646..14840)
FT      /rpt_family="LINE/L1"
FT      /rpt_unit-L1M4
FT      repeat_region      14866..15024      /rpt_family="DNA/MER1_type"
FT      /rpt_unit-MER33
FT      repeat_region      complement(15025..15166)
FT      /rpt_family="LINE/L1"
FT      /rpt_unit-L1MCB
FT      repeat_region      complement(15167..15478)
FT      /rpt_family="SINE/Alu"
FT      /rpt_unit-AluSg
FT      repeat_region      complement(15479..15491)
FT      /rpt_family="LINE/L1"
FT      /rpt_unit-L1MCB
FT      repeat_region      16180..16238      /rpt_family="LTR/MaLR"
FT      /rpt_unit-MTR1E2
FT      repeat_region      16288..16642      /rpt_family="LTR/MaLR"
FT      /rpt_family="LTR/MaLR"
FT      repeat_region      16666..17129      /rpt_family="LTR/MaLR"
FT      /rpt_unit-MTR1E2
FT      repeat_region      17198..17236      /rpt_family="Low_complexity"
FT      /rpt_unit-AT-rich
FT      repeat_region      complement(18160..18573)
FT      /rpt_family="LINE/L2"
FT      /rpt_unit-L2
FT      STS
FT      18303..18735
FT      /standard_name="SHGC-82788"
FT      19992..20045
FT      repeat_region      /rpt_family="SINE/MIR"
FT      /rpt_unit-MIR
FT      repeat_region      complement(20515..20823)
FT      /rpt_family="SINE/Alu"
FT      /rpt_unit-AluSg
FT      repeat_region      complement(20835..21083)
FT      /rpt_family="LINE/L1"

FT      repeat_region      /rpt_unit-L1PA8
FT      21753..22022      /rpt_family="SINE/Alu"
FT      /rpt_unit-AluSg
FT      repeat_region      22023..22053      /rpt_family="Low_complexity"
FT      /rpt_unit-AT-rich
FT      repeat_region      22196..22643      /rpt_family="LINE/L1"
FT      /rpt_unit-L1PA8
FT      repeat_region      complement(23919..23963)
FT      /rpt_family="LINE/L2"
FT      /rpt_unit-L2
FT      repeat_region      24040..24249      /rpt_family="SINE/MIR"
FT      /rpt_unit-MIR
FT      repeat_region      24302..24801      /rpt_family="SINE/Alu"
FT      /rpt_unit-AluSx
FT      repeat_region      complement(24859..24984)
FT      /rpt_family="LINE/L2"
FT      /rpt_unit-L2
FT      repeat_region      25258..25618      /rpt_family="LTR/MaLR"
FT      /rpt_unit-THE1B
FT      repeat_region      25849..26929      /rpt_family="LINE/L1"
FT      /rpt_unit-L1PA12
FT      repeat_region      26974..27007      /rpt_family="Low_complexity"
FT      /rpt_unit-AT-rich
FT      repeat_region      27907..28148      /rpt_family="SINE/Alu"
FT      /rpt_unit-AluSx
FT      repeat_region      28152..28179      /rpt_family="Low_complexity"
FT      /rpt_unit-AT-rich
FT      repeat_region      28669..28779      /rpt_family="LINE/L1"
FT      /rpt_unit-L1MEC
FT      repeat_region      29051..29290      /rpt_family="LINE/L1"
FT      /rpt_unit-L1MEC
FT      repeat_region      29342..30136      /rpt_family="LINE/L1"
FT      /rpt_family="Simple_repeat"
FT      30160..30331      /rpt_unit-(TA)n
FT      repeat_region      30339..30487      /rpt_family="LINE/L1"
FT      /rpt_unit-L1MEC
FT      repeat_region      30532..30819      /rpt_family="SINE/Alu"
FT      /rpt_unit-AluSx
FT      repeat_region      30901..31101      /rpt_family="LINE/L1"
FT      /rpt_unit-L1M4
FT      repeat_region      31102..31410

Query Match      18.1%; Score 34.8; DB 17; Length 166156;
Best Local Similarity 65.4%; Pred.No.3.8;
Matches 51; Conservative 0; Mismatches 27; Indels 0; Gaps 0;
```

RESULT	6	AC024216/c	LOCUS	AC024216	181870 bp	DNA	HTG	25-JUL-2000
DEFINITION		Human sapiens chromosome 3 clone RP11-204C23, WORKING. DRAFT						
ACCESSION		AC024216						
VERSION		AC024216.16		GI:9438304				
KEYWORDS		HTG: HTGS_PHASE1; HTGS_DRAFT.						
SOURCE		human.						
ORGANISM		Homo sapiens						
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.						
AUTHORS		Muzny D.M., Adams C., Bailey M., Barbara J., Blankenburg K., Bodor B., Buck J.J., Boyle S., Brooks A., Bulay C., Bunac C., Burkett C., Burrows J., Carter M., Chacho J., Chen Z., Cox C., David R., Delgado O., Deshazo D., Ding Y., Domah-Rashid N., Dugan-Rocha S., Durbin K.J., Fernandez C., Ferraguto D., Forcum-Tansey J., Frantz P., Ganesh R., Gorrell J.H., Gorrell L.L., Guevara W., Harris K., Hernandez J., Hodgson A., Hogues M., Holloway C., Hosak H., Jackson L.E., Jackson L., Jia Y., Jones M., Kelly S., Kondajewski N., Kong Y., Kovar C., Leal B., Li Z., Lichtarge O., Liu J., Liu W., Logan O., Lozado R.J., Lu J., Lucier R., Martin R., Martinez C., McLeod M.P., Mei G., Morgan M., Morris S., Nash S., Nelson A., Nguyen R., Nguyen N., Nguyen S., Owsal G., Parish B., Paxton S., Payton B., Perez L., Pu L.L., Quiles M., Reiter D., Rives M., Samuel S., Say J., Scherer S., Shah E., Shen H., Simon M., Sparks A., Stamps A., Succang R., Tabori P., Taylor T., Vasquez L., Vinson R., Vo O., Wahab M., Wallington S., Weinstein G., Weinstein I.R., Williamson A., Worley K., Wren J., Wrensford G., Yu W., Zhou X., Nelson D. and Gibbs R.						
TITLE		Direct Submission						
JOURNAL		Unpublished						
REFERENCE		2 (bases 1 to 181870)						
AUTHORS		Worley K.C.						
TITLE		Direct Submission						
JOURNAL		Submitted (28-FEB-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA						
COMMENT		On Jul 25, 2000 this sequence version replaced gi:9255957.						

*	63363	63462:	gap of unknown length
*	63463	78902:	contig of 15440 bp in length
*	78903	79002:	gap of unknown length
*	79003	89872:	contig of 10871 bp in length
*	89873	89973:	gap of unknown length
*	89974	97597:	contig of 7624 bp in length
*	97598	97697:	gap of unknown length
*	97698	105158:	contig of 7461 bp in length
*	105159	105258:	gap of unknown length
*	105259	112486:	contig of 7228 bp in length
*	112487	112586:	gap of unknown length
*	112587	117611:	contig of 5031 bp in length
*	117618	123759:	contig of 6042 bp in length
*	123760	123859:	gap of unknown length
*	123860	128429:	contig of 4570 bp in length
*	128430	128529:	gap of unknown length
*	128530	133354:	contig of 3825 bp in length
*	132355	132454:	gap of unknown length
*	132455	135891:	contig of 3437 bp in length
*	135892	135991:	gap of unknown length
*	135992	141318:	contig of 5327 bp in length
*	141319	141418:	gap of unknown length
*	141419	145280:	contig of 3862 bp in length
*	145281	145380:	gap of unknown length
*	145381	148462:	contig of 3082 bp in length
*	148463	148562:	gap of unknown length
*	148563	151476:	contig of 2914 bp in length
*	151477	151576:	gap of unknown length
*	151577	154759:	contig of 3183 bp in length
*	154760	154859:	gap of unknown length
*	154860	156422:	contig of 1566 bp in length
*	156426	156525:	gap of unknown length
*	156526	158715:	contig of 2190 bp in length
*	158716	158815:	gap of unknown length
*	158816	161739:	contig of 2924 bp in length
*	161740	161839:	gap of unknown length
*	161840	163737:	contig of 1534 bp in length
*	163744	163773:	gap of unknown length
*	163474	164495:	contig of 1022 bp in length
*	164496	164595:	gap of unknown length
*	164596	166256:	contig of 1661 bp in length
*	166257	166356:	gap of unknown length
*	166357	168039:	contig of 1683 bp in length
*	168040	168139:	gap of unknown length
*	168140	169184:	contig of 1045 bp in length
*	169185	169284:	gap of unknown length
*	169285	170463:	contig of 1179 bp in length
*	170464	170563:	gap of unknown length
*	170564	172541:	contig of 1978 bp in length
*	172542	172641:	gap of unknown length
*	172642	174280:	contig of 1639 bp in length
*	174281	174380:	gap of unknown length
*	174381	175663:	contig of 1283 bp in length
*	175664	175763:	gap of unknown length
*	175764	176997:	contig of 1234 bp in length
*	176998	177097:	gap of unknown length
*	177098	178231:	contig of 1134 bp in length
*	178232	178331:	gap of unknown length
*	178332	179583:	contig of 1252 bp in length
*	179584	179683:	gap of unknown length
*	179684	180695:	contig of 1012 bp in length
*	180696	180795:	gap of unknown length
*	180796	181870:	contig of 1075 bp in length.

Location/unaffliers
1. -181870

```

"/organism="Homo sapiens"  
/db_xref="taxon:9606"  
/chromosome="3"  
/clone="RE11-204C23"  


```

BASE COUNT 53768 a 34063 c 34457 g 56115 t 3467 others

ORIGIN

Query Match	18.1%;	Score 34.8;	DB 2;	Length 181870;
Best Local Similarity	65.4%;	Pred. No. 3.9;		
Matches 51;	Conservative 0;	Mismatches 27;	Indels 0;	Gaps 0;

```

Oy  47  tgaatcgtccatcttccaataccctcggggaatatagtgtaaaaccctataaaacgcyg  106
      |||  |  |||||  |  |||||  |  |||||  |  |||||  |  |||||  |  |||||  |
Db  21069  TGAAGAAGACTATTTTGTCACACCTTTGTGCAATATATTTGAAATCTAGATGAATACAA  21010

```

```

Oy      107      gtttcgcagaacatgc      124
          ||||  ||||  |||
Db      21009    TATTCTAGAAATAATAC      20992

```

	RESULT	7
AC018497		
LOCUS		
DEFINITION	AC018497 188432 bp DNA	HTG 03-FEB-2000
DESCRIPTION	Homo sapiens chromosome 3p clone RP11-229E21, WORKING DRAFT	
SEQUENCE	SEQUENCE, 11 unordered pieces.	
AC018497		
AC018497		

VERSION	AC018497.4	GI:6862671
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT	
SOURCE	human.	
ORGANISM	Homo sapiens	

REFERENCE
AUTHORS

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 18432)
Wu, Q., Hu, S., Dong, W., Zhang, X., Wang, J., Zhang, Y., Zhang, H.,

TITLE	Chromosome 3p genomic sequence
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 188432)
AUTHORS	Bao, Q., Hu, S., Dong, W., Wang, J., Zhang, Y., Zhang, H., Liu, B.,

TITLE Direct Submission
JOURNAL Submitted (13-DEC-1999) Human Genomic Center, Institute of Genetics, Chinese Academy of Sciences, Datun Road, Beijing 100101, P.R.China
COMMENT On Feb 3, 2000 this sequence.version replaced q1:6684200.

On Feb 3 2000 this sequence version replaced gi.1.6664200
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 11 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

*	1	2725:	contig of 2725 bp in length
*			gap of unknown length
*	2726	4464:	contig of 1739 bp in length
*			gap of unknown length
*	4465	5842:	contig of 1378 bp in length
*			gap of unknown length
*	5843	9661:	contig of 3819 bp in length
*			gap of unknown length
*	9662	16331:	contig of 6670 bp in length
*			gap of unknown length
*	16332	24690:	contig of 8359 bp in length
*			gap of unknown length
*	24691	40720:	contig of 16030 bp in length
*			gap of unknown length
*	40721	72584:	contig of 31864 bp in length
*			gap of unknown length
*	72585	106317:	contig of 33733 bp in length
*			gap of unknown length

FEATURES	
*	106318 142895: contig of 36578 bp in length
*	gap of unknown length
*	142896 188432: contig of 45537 bp in length
	Location/Qualifiers

BASE COUNT	a	c	g	t	other
ORIGIN	60668	35322	35795	56644	3

Query Match	18.1%	Score 34.8	DB 2	Length 188432
Best Local Similarity	65.4%	Pred. No. 3.9		
Matches	51	Conservative	0	Mismatches 27
				Indels 0
				Gaps 0

0y 47 tgcgtcgctatatttcacatccctcggggaatagatcgtgaaacccataaaacgcgg 106
 ||| | ||||| | ||||| | ||||| | ||||| | ||||| |
 Db 133724 TGAAGAAGACTATTTTGTACACCTTTGTGCAATATATTTGAAAAATCTAGATGAAATACAA 133785

```

ay      10/  gtttcgcgaacatgc 124
          ||| ||| ||| |||
Db 133784 TATTTCTTAGAAAAATAC 133801

```

RESULT	8		
AC026486			
LOCUS	AC026486	148259 bp	DNA
DEFINITION	Homo sapiens clone RPL1-2AK7, WORKING DRAFT SEQUENCE, 18 unordered clones		HTG 06-MAY-2001

```

VERSION      AC026486.2  GI:9954792
KEYWORDS     HTG; HTGS_PHASE1; HTGS_DRAFT
SOURCE       human.

```

REFERENCE
1 (bases 1 to 148269)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

TITLE	Homo sapiens, clone RP11-24K/
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 148269)
AUTHORS	Birren B., Jinton L., Nusbaum C., Lander E., Abraham H., Allen N.

Boguslavsky, L., Boukhgalter, B., Brown, A., Burnett, G., Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Deatellano, K., Dewar, K., Diaz, J.S.,

Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Hotton, L., Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.

Levine, K., Liu, C., Liu, G., Locke, K., MacDonald, P., Marquis, N., McCarthy, M., McEwan, P., McGuck, A., McKernan, K., McPheeters, R., Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Murphy, T., Navlor, J., Norman, C. H., O'Connor, T., O'Donnell, B.

Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,

TITLE
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.-J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Submitted 22 Jun 2000; published 14 Jul 2000

COMMENT
RESEARCH, 320 CHAIRES STREET, CAMBRIDGE, MA 02141, USA
On Aug 31, 2000 this sequence version replaced gi:7280316.
All repeats were identified using RepeatMasker:
Smith, A.F. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research

```

Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information -----
Center project name: 14489
Center clone name: 24_K7
----- Summary Statistics -----
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 140190 bases at least Q40
Consensus quality: 144274 bases at least Q30
Consensus quality: 145694 bases at least Q20
Insert size: 145000; agarose-1p
Insert size: 146569; sum-of-contigs
Quality coverage: 4.6 in Q20 bases; agarose-1p
Quality coverage: 4.5 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1. 1564: contig of 1564 bp in length
* 1565 1664: gap of 100 bp
* 1665 20700: contig of 19036 bp in length
* 20701 20800: gap of 100 bp
* 20801 21849: contig of 1049 bp in length
* 21850 21949: gap of 100 bp
* 21950 23604: contig of 3655 bp in length
* 23605 25704: gap of 100 bp
* 25705 29764: contig of 4060 bp in length
* 29765 29864: gap of 100 bp
* 29865 35579: contig of 5715 bp in length
* 35580 35679: gap of 100 bp
* 35680 41577: contig of 5898 bp in length
* 41578 41677: gap of 100 bp
* 41678 47413: contig of 5736 bp in length
* 47414 47513: gap of 100 bp
* 47514 54860: contig of 7347 bp in length
* 54861 54960: gap of 100 bp
* 54961 61589: contig of 6629 bp in length
* 61590 61689: gap of 100 bp
* 61690 68215: contig of 6526 bp in length
* 68216 68315: gap of 100 bp
* 68316 78737: contig of 10422 bp in length
* 78738 78837: gap of 100 bp
* 78838 89917: contig of 11080 bp in length
* 89918 90017: gap of 100 bp
* 90018 100805: contig of 10788 bp in length
* 100806 100905: gap of 100 bp
* 100906 114689: contig of 13784 bp in length
* 114690 114789: gap of 100 bp
* 114790 132632: contig of 17843 bp in length
* 132633 132732: gap of 100 bp
* 132733 147978: contig of 15246 bp in length
* 147979 148078: gap of 100 bp
* 148079 148269: contig of 191 bp in length.
Location/Qualifiers
1.148269
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-24K7"
/clone_1lb="RPC1-11 Human Male BAC"
1.1564
/note="assembly-fragment"
misc_feature
1.1564
/note="assembly-fragment"
clone_end:SP6
vector_side:left"
misc_feature
1665..20700
/note="assembly-fragment"

```

```

misc_feature 20801..21849 /note="assembly-fragment"
misc_feature 21950..25604 /note="assembly-fragment"
misc_feature 25705..29764 /note="assembly-fragment"
misc_feature 29865..35579 /note="assembly-fragment"
misc_feature 35680..41577 /note="assembly-fragment"
misc_feature 41678..47413 /note="assembly-fragment"
misc_feature 47514..54860 /note="assembly-fragment"
misc_feature 54961..61589 /note="assembly-fragment"
misc_feature 61690..68215 /note="assembly-fragment"
misc_feature 68316..78737 /note="assembly-fragment"
misc_feature 78838..89917 /note="assembly-fragment"
misc_feature 90018..100805 /note="assembly-fragment"
misc_feature 100906..114689 /note="assembly-fragment"
misc_feature 114790..132632 /note="assembly-fragment"
misc_feature 132733..147978 /note="assembly-fragment"
misc_feature 148079..148269 /note="assembly-fragment"
misc_feature 148269..148269 /note="assembly-fragment"
BASE COUNT 46902 a 26848 c 26913 g 45901 t 1705 others
ORIGIN
Query Match 17.9%; Score 34.4; DB 2; Length 148269;
Best Local Similarity 52.0%; Pred. No. 5;
Matches 77; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
QY 27 ttgcaagcatatcttgcctgctcatttccttcgaggaataagatgtg 86
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 140410 ATTTCATATTATATGTTTCATTGATGATGATTCACACATTGCTTCATATTG 140469
QY 87 aaacctataaacgcggttttcgcagaacacatgcctagatcatcattgacaca 146
| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 140470 ATTTCATATTATGATTCATGATGATTCATGATTCATGATTCATGATTCAT 140529
QY 147 tggactaagcaaaagtgcctgcctcctg 174
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 140530 TGTACTAGAGAAATTCCTTCCTTGG 140557
RESULT 9
AC020682 AC020682 162249 bp DNA HTG 19-JUL-2000
LOCUS Homo sapiens clone RP11-24G18, WORKING DRAFT SEQUENCE, 13 unordered
pieces.
ACCESSION AC020682
VERSION AC020682.3 GI:9280776
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 162249)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome, clone RP11-24G18
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 162249)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

```


AUTHORS Wray, P.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquay@sanger.ac.uk

COMMENT

On May 11, 2001 this sequence version replaced g1:11995192. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPep; Information on the WormPep database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr10>

RP11-63A2 is from the library RP11-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/Bacpac/home.htm>

VECTOR: PBAC3.6

This sequence is the entire insert of clone RP11-63A2 The true right end of clone RP11-809M12 is at 46921 in this sequence.

FEATURES

source

Location/Qualifiers

1. 166973
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="10"
 /clone="RP11-63A2"
 /clone_11b="RP11-11.1"
 3. 805
 /note="LMC5 repeat: matches 7127. .7913 of consensus"
 1250. 1679
 /note="MLT2B repeat: matches 1. .444 of consensus"
 1882. 4635
 /note="HERVL repeat: matches 2635. .5652 of consensus"
 4636. 4874
 /note="AluSq repeat: matches 67. .297 of consensus"
 4875. 7529
 /note="HERVL repeat: matches 1. .2635 of consensus"
 7530. 7577
 /note="MLT2A repeat: matches 406. .453 of consensus"
 7594. 8007
 /note="MLT2A repeat: matches 150. .450 of consensus"
 8233. 8517
 /note="MUT1-INTERNAL repeat: matches 932. .1211 of consensus"
 8555. 8854
 /note="LMC/D repeat: matches 5313. .5604 of consensus"
 8913. 9043
 /note="LMC/D repeat: matches 5604. .5735 of consensus"
 9108. 9312
 /note="LMC5 repeat: matches 7273. .7456 of consensus"
 9469. 9845
 /note="LMC5 repeat: matches 7507. .7895 of consensus"
 10014. 10105
 /note="46 copies 2 mer aa 63% conserved"
 10601. 10716
 /note="LM4 repeat: matches 2695. .2818 of consensus"
 10829. 11016
 /note="LM4 repeat: matches 2957. .3146 of consensus"
 11333. 11390

repeat_region
 11411. 11516
 /note="MLT2D repeat: matches 1. .58 of consensus"
 repeat_region
 11612. 11968
 /note="LM4 repeat: matches 2957. .3062 of consensus"
 repeat_region
 11986. 12049
 /note="MLT2D repeat: matches 70. .435 of consensus"
 repeat_region
 12683. 12935
 /note="32 copies 2 mer aa 67% conserved"
 repeat_region
 13751. 14135
 /note="LM3A repeat: matches 5659. .5936 of consensus"
 repeat_region
 14156. 14460
 /note="LM repeat: matches 471. .848 of consensus"
 repeat_region
 14461. 14976
 /note="Aluub repeat: matches 1. .300 of consensus"
 repeat_region
 14661. 14976
 /note="LM repeat: matches 848. .1572 of consensus"
 repeat_region
 17233. 17352
 /note="3 copies 40 mer 75% conserved"
 repeat_region
 17242. 17345
 /note="26 copies 4 mer gaag 72% conserved"
 repeat_region
 17276. 17359
 /note="14 copies 6 mer agaag 65% conserved"
 repeat_region
 18041. 18273
 /note="MIR repeat: matches 8. .246 of consensus"
 repeat_region
 19550. 19633
 /note="42 copies 2 mer ta 83% conserved"
 repeat_region
 19554. 19633
 /note="2 copies 40 mer 91% conserved"
 repeat_region
 20024. 22103
 /note="LIPB3 repeat: matches 4126. .6150 of consensus"
 repeat_region
 26100. 26159
 /note="35 copies 2 mer at 77% conserved"
 repeat_region
 26101. 26164
 /note="16 copies 4 mer tata 79% conserved"
 repeat_region
 26106. 26165
 /note="5 copies 12 mer 81% conserved"
 repeat_region
 26204. 27088
 /note="LMC3 repeat: matches 6716. .7586 of consensus"
 repeat_region
 27760. 27918
 /note="MLT13 repeat: matches 201. .346 of consensus"
 repeat_region
 28811. 28944
 /note="MIR repeat: matches 12. .149 of consensus"
 repeat_region
 29062. 29149
 /note="MLT1D repeat: matches 2. .89 of consensus"
 repeat_region
 30829. 30884
 /note="MIR repeat: matches 76. .131 of consensus"
 repeat_region
 33710. 33949
 /note="MIR repeat: matches 13. .262 of consensus"
 repeat_region
 34018. 34106
 /note="12 repeat: matches 2655. .2747 of consensus"
 repeat_region
 34360. 34452
 /note="MIR repeat: matches 50. .144 of consensus"
 repeat_region
 35653. 35789
 /note="MIR repeat: matches 3. .137 of consensus"
 repeat_region
 36130. 36201
 /note="12 copies 6 mer atatat 69% conserved"
 repeat_region
 36131. 36202
 /note="36 copies 2 mer ta 70% conserved"
 repeat_region
 37945. 37968
 /note="12 copies 2 mer tc 100% conserved"
 repeat_region
 38866. 39253
 /note="LM3A repeat: matches 5912. .6304 of consensus"
 repeat_region
 39256. 39337
 /note="THEIR repeat: matches 1. .82 of consensus"
 repeat_region
 39340. 41088
 /note="LM1 repeat: matches 971. .2727 of consensus"
 repeat_region
 41115. 41324
 /note="LMC5 repeat: matches 7701. .7912 of consensus"
 repeat_region
 41359. 41767
 /note="LM8 repeat: matches 5725. .6145 of consensus"
 repeat_region
 41771. 41943
 /note="LMC5 repeat: matches 7603. .7774 of consensus"
 repeat_region
 43197. 43436
 /note="MIR repeat: matches 16. .260 of consensus"

```
repeat_region 43580..43882
/note="AlusC repeat: matches 1..305 of consensus"
repeat_region 44598..44892
/note="Alusq repeat: matches 1..296 of consensus"
repeat_region 45115..45156
/note="LTR16B repeat: matches 113..160 of consensus"
repeat_region 49483..49604
/note="MIR repeat: matches 141..262 of consensus"
repeat_region 49614..49668
/note="L2 repeat: matches 2626..2682 of consensus"
repeat_region 50064..50218
/note="MIR repeat: matches 12..197 of consensus"
repeat_region 50244..50526
/note="AlusX repeat: matches 1..286 of consensus"
repeat_region 51305..51539
/note="L1PA16 repeat: matches 5913..6157 of consensus"
repeat_region 52439..52618
/note="MER58A repeat: matches 19..224 of consensus"
repeat_region 52647..52871
/note="MIR repeat: matches 2..230 of consensus"
repeat_region 54233..54312
/note="2 copies 40 mer 95% conserved"
repeat_region 55178..55369
/note="LTR1A1 repeat: matches 1..180 of consensus"
repeat_region 55428..55613
/note="LTR1A1 repeat: matches 181..374 of consensus"
repeat_region 55992..56287
/note="AlusX repeat: matches 3..307 of consensus"
repeat_region 57708..57735
/note="14 copies 2 mer aa 100% conserved"
repeat_region 59512..59581
/note="35 copies 2 mer tc 92% conserved"
repeat_region 59514..59581
/note="17 copies 4 mer tctc 94% conserved"
repeat_region 60185..60295
/note="MIR repeat: matches 6..120 of consensus"
repeat_region 61009..61056
/note="8 copies 6 mer tacaca 79% conserved"
repeat_region 61011..61054
/note="11 copies 4 mer caca 81% conserved"
repeat_region 61281..62062
/note="L2 repeat: matches 1849..2714 of consensus"
repeat_region 62072..62276
```

Query Match 17.9%; Score 34.4; DB 9; Length 166973;
Best Local Similarity 52.0%; Pred. No. 5.1;
Matches 77; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

```
Qy 27 ttgacagcataactctgtcgtatcttcaataacttcggggaatagatgtg 86
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 38795 TTATCAATTATATGTTTCAATGATATGTTGATCAAGACATTTGCTTCATATATTGG 38736

Qy 87 aaacccctataaaacgcgggtttcgcagaacaacatgcgcagtagatcattgacaca 146
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 38735 ATTATCAATTAATGAATATCAAAATTTGTCATGCCAATTTATATATATGATCAATTGGA 38676

Qy 147 tggactaagcaaaagtctgtccctcg 174
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 38675 TGTACTAGAGCAAAATCTCTCTCCTTG 38648
```

RESULT 11
AL356952/c 188026 bp DNA HTG 21-JUL-2001
LOCUS Homo sapiens chromosome 10 clone RP11-809M12, *** SEQUENCING IN
DEFINITION PROGRESS ***, In ordered pieces.
ACCESSION AL356952
VERSION AL356952.17 GI:15020905
KEYWORDS HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 188026)
AUTHORS Blakey,S.
TITLE Direct Submission
JOURNAL Submitted (20-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Jul 25, 2001 this sequence version replaced gi:14970342.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: BA809M12
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big dye; 2% of reads
Chemistry: Dye-primer Big dye; 97% of reads
Consensus quality: 187552 bases at least Q40
Consensus quality: 187814 bases at least Q30
Consensus quality: 187920 bases at least Q20
Insert size: 188026; sum-of-contigs
Insert size: 167487; 12.8% error; agarose-fp
Quality coverage: 6.20x in Q20 bases; sum-of-contigs quality
coverage: 6.96x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source
1..188026
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-809M12"
/clone_id="RPC1-11.3"
1..188026
/note="assembly fragment:01381"

misc_feature
BASE COUNT 61435 a 34467 c 34626 g 57498 t
ORIGIN

Query Match 17.9%; Score 34.4; DB 2; Length 188026;
Best Local Similarity 52.0%; Pred. No. 5.2;
Matches 77; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

```
Qy 27 ttgacagcataactctgtcgtatcttcaataacttcggggaatagatgtg 86
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 179899 TTATCAATTATATGTTTCAATGATATGTTGATCAAGACATTTGCTTCATATATTGG 179840

Qy 87 aaacccctataaaacgcgggtttcgcagaacaacatgcgcagtagatcattgacaca 146
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 179839 ATTATCAATTAATGAATATCAAAATTTGTCATGCCAATTTATATATATGATCAATTGGA 179780

Qy 147 tggactaagcaaaagtctgtccctcg 174
|| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 179779 TGTACTAGAGCAAAATCTCTCTCCTTG 179752
```

RESULT 12
VYIVD10/c 155711 bp DNA INV 06-JUL-2000
LOCUS Plasmodium vivax telomeric YAC clone, complete finished sequence.
DEFINITION ACCESION A3360354
VERSION A3360354.1 GI:8935641
KEYWORDS malaria parasite P. vivax.
SOURCE Plasmodium vivax
ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 155711)
AUTHORS Oliver,K., Bowman,S., Hall,N., Quail,M., Rajandream,M.A.,

TITLE Harris, D., del Portillo, H.A., Lanzer, M. and Barrell, B.G.
JOURNAL Direct Submission
Submitted (27-JUN-2000) The Sanger Centre, Wellcome Trust Genome
Campus, Hinxton, Cambridge CB10 1SA, UK
COMMENT For more information about this project
see http://www.sanger.ac.uk/Projects/P_vivax.
FEATURES
source
1. 155711
/organism="Plasmodium vivax"
/db_xref="taxon:5855"
/country="Brazil"
repeat_region
gene complement(join(1173..1337,1442..2149,2276..2302))
1. 611
/gene="v1r1"
complement(join(1173..1337,1442..2149,2276..2302))
CDS
/gene="v1r1"
/note="v1r1, rpf3, putative transmembrane protein, similar
to v1r33: MW:35552 (301 aa) fasta scores: opt: 437, E():
0.00013, 26.5% identity in 302 aa overlap"
/codon_start=1
/protein_id="CAB96690.1"
/db_xref="GI:8953642"
/translation="MADEYNDSEKRYVVISFPKNKEYESVTPSSDRTIVSACNOIRSK
ISAFSPYTCVKTQYLVHINHEYANODVNDKCEFLVWINSINVINPQVDTSTF
FNDMSVYKEOKSKINICINNMKYNINNVLDKILMDLFGNPKKEIKNEKEDC
SLGECVSVYIGSLDKCKDNNTKRCNILEVFNKYNIEAPRIGYCNVORLYPVKG
VLNVSAYIANFTPIILSSYIFELFKETPILOSMTROPQVEKKILKLNLOETFEKLOE
NKNIESGTRNGENFLAYNAVN"
/join(7387..7404,7544..8248,8372..8527)
/gene="v1r2"
/join(7387..7404,7544..8248,8372..8527)
/gene="v1r2"
/note="v1r2, putative transmembrane protein, similar to
v1r13: MW:36805 (312 aa) fasta scores: opt: 380, E():
0.00049, 28.0% identity in 296 aa overlap, and to v1r9:
MW:34499 (294 aa), fasta scores: opt: 377, E(): 0.00053,
27.2% identity in 290 aa overlap"
/codon_start=1
/protein_id="CAB96691.1"
/db_xref="GI:8953643"
/translation="MGCESEMHFNSRSEHESLPLNTNKNVQGHDIKCEKMSAYSI
YKSTFNLCNKSLNLEYLSEHEDIRKSGTGLVLMYKRELKNVKNKSHLDY
NNLLEICEFYISYNTIGTYSQNVNRADNFEILKLDLYKPKIKDYDCENTYKCA
KNCPDYKYEIDCKKKNRDPNCNLENFNPNFNPYISSEPECKKDLIPIKNTYTK
SVPLISLVTVLSVPFFFLIKLTPFGAMFNRDKSGNKNKNSNIAEPEPLHASRI
SRMSDIKTKYISYMSK"
/join(11806..11817,11933..12763,12882..13034)
/gene="v1r3"
/join(11806..11817,11933..12763,12882..13034)
/gene="v1r3"
/note="v1r3, rpf1, putative transmembrane protein, member
of rpa subfamily, similar to rpa2, pseudogene 5', v1r31:
MW:26193 (221 aa) fasta scores: opt: 862, E(): 8.2e-16,
55.5% identity in 218 aa overlap and rpa2, pseudogene 3',
and to rpa3, v1r17 MW:46468 (406 aa) fasta scores: opt:
415, E(): 1.9e-06, 27.7% identity in 393 aa overlap"
/codon_start=1
/protein_id="CAB96692.1"
/db_xref="GI:8953644"
/translation="MTSISNFEDECKLSAFIKSLDSSPOMQAVDIDSYLNSKREOKD
TVREIYSELESYSIASROEDNIKSGCFYLMYLNKORKEKLIWNVNDAMVITKLM
GKTRPSVSCROHAYEPESEKRCVDMAYCVNDELOKHEKPDQAEKSTYDNE
NKYTRHYTYFTTNKCLRDTRNDIHPMKRSDDTCLHDMATKPEKFTESSTIYDGG
TREKTRKESHEHDSRINCVMIDGVPLLEISATINVIPLKYGIIAGSSPFGTISLG
LYLKTRTHPSLVTRNSSRENKINNTDKLSHEKESKSNKNDYKFSYNDION"
/join(16650..16682,16811..17680,17800..17964)
/gene="v1r4"
/join(16650..16682,16811..17680,17800..17964)
/gene="v1r4"
/note="v1r4, rpf2, putative transmembrane protein, member of
rpf subfamily, similar to rpf3, v1r14: MW:40917 (348 aa)
fasta scores: opt: 878, E(): 8.4e-16, 42.1% identity in
349 aa overlap, and to rpf1, v1r10: 0.4: v1r10: 0.4: tab

12087:13361 reverse MW:39223 (343 aa) fasta scores: opt:
597, E(): 5.7e-10, 44.7% identity in 351 aa overlap and to
rpf4, v1r10: 0.20: v1r10: 0.20: tab 81364:82388 reverse MW:34896
(299 aa) fasta scores: opt: 441, E(): 8.7e-07, 35.4%
identity in 339 aa overlap"
/codon_start=1
/protein_id="CAB96693.1"
/db_xref="GI:8953645"
/translation="MTEKDYPGLISVETIISDPSNKFYNYLNNYDNLKKYRYCT
SLNPKSKDITNTNLICAOVLYKTYTTLNKDGLYDDCLILNTVYISRLVNLIG
TEDKTVIAPIRGLVWINSIVANPLVNTSKPYNKNDKIIPTONDNRKRELYDCNV
YDTIKNTLPYEDOMCEPKYMYNVESTSLFEKFTLCSKSDQCEPEFYECNOYDNPV
LTFECNETMFKKAKKEETSEKDKLOLQSGOGEANGMGTNPEFSGSHLPDGT
HPAOTGNIILGVATSTISGALVRFPLGMNLNRGFRNNHNNRMHMGEGYGLDYA
PESTNPTPGGGEHETGHPA"
/join(20974..21000,21140..21889,22031..22477,22627..22788)
/gene="v1r5"
/join(20974..21000,21140..21889,22031..22477,22627..22788)
/gene="v1r5"
/note="v1r5, hypothetical protein, splicing uncertain,
low similarity to other v1r10 CDS"
/codon_start=1
/protein_id="CAB96694.1"
/db_xref="GI:8953646"
/translation="MEEVEERKDYDFKLDLDYETIIHQVNRYSDEPOANCDISITF
TNEKNQVSVLCRKVYSFEEKIKHNRYSNPODKKYPDLNFMIRHLOLOSISKY
DSVYIHLKNKNYPODKERELEENKIYINERDKSNMILDLKNTYGSLEHETEC
NYFOSFEKENDKLCYRCYFGHDSKLCVNMIFELVDKOFPNONCHITSTYSLP
ELSDSSEKTKTDSKDEPGLKVTSTKSPFQIQAIVYKELNELIRQYNNPY
YDDEKMKFMKMLILQFIKCYRENKKNKLSFPMKEFTEKYKKNENYNDIADACKI
GCKSKIYGLKSKCENYGDLLILEKVNYSYLEOKKYISLSLEWITKRAOSLFO
DGMASRSPVVISMTPLPSISIRKEKKRKIPLEFPERSINDYKNNSNMKMKSK
RKRIRATQPN"
/join(24943..24969,25109..26278,26390..26551)
/gene="v1r6"
/join(24943..24969,25109..26278,26390..26551)
/gene="v1r6"
/note="v1r6, putative transmembrane protein, similar to
v1r28: (458 aa), fasta scores: opt: 344, E(): 0.00048,
26.3% identity in 468 aa overlap"
/codon_start=1
/protein_id="CAB96695.1"
/db_xref="GI:8953647"
/translation="MAGDHOPGYISYNDVYNAKHEERFRYNNRPNDREREKTIINKI
NNSHKMMLKKTFTTLNVLANDLAELFGKMHCRITNWLKNEVANNHNPDKSY
EPYFOKFSDESIITKTNNDORCNNTYINLHKITINMIDLYGLYDEDKIKSHRDS
NYGSDPFLSWASNNHVAIDRYENTNLYOFEEIEKILIDMLKESSSSCIKSYLAK
POVYKLRSEERARREAAKORALEAERREERARROHDELAERELDROANVT
GDTASGEGTPOHTDRDLRAPITISELDONNORGVPRSSRGLIDNCHTRGCV
ITYNEDTRQGEYVIEDGTIKTSNGMGSSGSPRYITVEFGSDYDPVYVSGGMC
ALFLFRTPLDTEFFRGGRAHRIPRSFNQLGAPFDINENGYIGYPMIDPYG
AE"
/join(29296..29313,29447..30361,30480..30638)
/gene="v1r7"
/join(29296..29313,29447..30361,30480..30638)
/gene="v1r7"
/note="v1r7, rpf1, putative transmembrane protein, member
of rpf subfamily, similar to rpf2, v1r29: MW:42155 (362
aa) fasta scores: opt: 805, E(): 1.4e-07, 39.8% identity
in 374 aa overlap, and to rpf3, v1r27: MW:42372 (362 aa)
fasta scores: opt: 487, E(): 0.0004, 26.9% identity in 376
aa overlap"
/codon_start=1
/protein_id="CAB96696.1"
/db_xref="GI:8953648"
/translation="MAEILGDEKLEILPTKNNYVLYDNGYFCEKDIYFNHAEKRLSS
DSGLQVYERKILKAMCYWYTKRKRIKIPESIDKCFEYVYLAILINLNDHFTSVYLI
NLYRLINAGAGKICDPIINSYIDNDNPEINLIDYSDYSYILALDAPRSGNSVLI
KKYIQTIVNKRKEALGCEVKEKSDTYCKEKKHPSKNEHMLSTWCSNLEGTQQLN
LIGDQGVVEEAQIGKSTGKGAELKEKTVBGRADPVLPLASLEGGHMLSS
SSEVEDSSISTTGYLANVASVAGIYPSLVYPAQWIMRLFGRTPEKMMNTL
ADMEILNFSOTGSEFNERSRFDISYRV"
/join(34090..34134,34277..35083,35199..35366)


```

* 7279 8126: contig of 848 bp in length
* 8127 8146: gap of unknown length
* 8147 9391: contig of 1245 bp in length
* 9392 9411: gap of unknown length
* 9412 10106: contig of 695 bp in length
* 10107 10126: gap of unknown length
* 10127 11256: contig of 1130 bp in length
* 11257 11276: gap of unknown length
* 11277 12281: contig of 1005 bp in length
* 12282 12301: gap of unknown length
* 12302 13305: contig of 1004 bp in length
* 13306 13325: gap of unknown length
* 13326 14042: contig of 717 bp in length
* 14043 14062: gap of unknown length
* 14063 14871: contig of 809 bp in length
* 14872 14891: gap of unknown length
* 14892 15691: contig of 800 bp in length
* 15692 15711: gap of unknown length
* 15712 17102: contig of 1391 bp in length
* 17103 17122: gap of unknown length
* 17123 18533: contig of 1411 bp in length
* 18534 18553: gap of unknown length
* 18554 19689: contig of 1136 bp in length
* 19690 19709: gap of unknown length
* 19710 21533: contig of 1824 bp in length
* 21534 21553: gap of unknown length
* 21554 23430: contig of 1877 bp in length
* 23431 23450: gap of unknown length
* 23451 25116: contig of 1666 bp in length
* 25117 25136: gap of unknown length
* 25137 26550: contig of 1414 bp in length
* 26551 26570: gap of unknown length
* 26571 28056: contig of 1486 bp in length
* 28057 28076: gap of unknown length
* 28077 28243: contig of 167 bp in length
* 28244 28263: gap of unknown length
* 30115 30116: contig of 1852 bp in length
* 30116 30135: gap of unknown length
* 30136 31923: contig of 1788 bp in length
* 31924 31943: gap of unknown length
* 31944 33125: contig of 1182 bp in length
* 33126 33145: gap of unknown length
* 33146 33955: contig of 810 bp in length
* 33956 33975: gap of unknown length
* 33976 35756: contig of 1781 bp in length
* 35757 35776: gap of unknown length
* 35777 37277: contig of 1501 bp in length
* 37278 37297: gap of unknown length
* 37298 38619: contig of 1322 bp in length
* 38620 38639: gap of unknown length
* 38640 39780: contig of 1141 bp in length
* 39781 39800: gap of unknown length
* 39801 41047: contig of 1247 bp in length
* 41048 41067: gap of unknown length
* 41068 41938: contig of 871 bp in length
* 41939 41958: gap of unknown length
* 41959 43179: contig of 1221 bp in length
* 43180 43199: gap of unknown length
* 43200 44738: contig of 1539 bp in length
* 44739 44758: gap of unknown length
* 44759 46101: contig of 1343 bp in length
* 46102 46121: gap of unknown length
* 46122 47436: contig of 1315 bp in length
* 47437 47456: gap of unknown length
* 47457 49356: contig of 1900 bp in length
* 49357 49376: gap of unknown length
* 49377 51129: contig of 1753 bp in length
* 51130 51149: gap of unknown length
* 51150 53630: contig of 2481 bp in length
* 53631 53650: gap of unknown length
* 53651 56012: contig of 2362 bp in length
* 56013 56032: gap of unknown length
* 56033 57861: contig of 1829 bp in length.

```

```

FEATURES
  source
    Location/Qualifiers
      1..57861
        /organism="Drosophila melanogaster"
        /db_xref="taxon:7227"
        /clone="RP8-24A2"
BASE COUNT  15582 a 12043 c 11927 g 15854 t 2455 others
ORIGIN
  Query Match 17.7%; Score 34; DB 9; Length 166872;
  Best Local Similarity 61.1%; Pred. No. 5.9;
  Matches 55; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
  Oy 3 tggcgccggcgccctgaatgtgcttgacaacatactcttcgactatcttc 62
  Db 44537 tggcgcttactgttaaccgacggcgacagatracattttttatgtctttttt 44596
  Oy 63 aataccctcgaggaaatagatgtgaacc 92
  Db 44597 gattaccttgggaaatagcttttacttcc 44626

RESULT 14
AC008554/c 166872 bp DNA PRI 01-JUL-2000
LOCUS Homo sapiens chromosome 19 clone CTC-513N18, complete sequence.
DEFINITION AC008554
ACCESSION AC008554
VERSION AC008554.7 GI:8886966
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 166872)
  DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE Direct Submission
  JOURNAL Unpublished
  AUTHORS 2 (bases 1 to 166872)
  TITLE DOE Joint Genome Institute.
  JOURNAL Direct Submission
  AUTHORS 3 (bases 1 to 166872)
  TITLE DOE Joint Genome Institute and Stanford Human Genome Center.
  JOURNAL Submitted (01-JUL-2000) DOE Joint Genome Institute, 2800 Mitchell
  Drive, Walnut Creek, CA 94598, USA
  On Jul 1, 2000 this sequence version replaced g1:7711296.
  Draft Sequence Produced by DOE Joint Genome Institute
  WWW: jgi.doe.gov
  Finishing Completed at Stanford Human Genome Center
  www.shgc.stanford.edu
  Quality: Phrap Quality >=40 99.7% of Sequence;
  Estimated Total Number of Errors is 2.1.
FEATURES
  source
    Location/Qualifiers
      1..166872
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="19"
        /clone="CTC-513N18"
BASE COUNT  50087 a 32869 c 34967 g 48949 t
ORIGIN
  Query Match 17.7%; Score 34; DB 9; Length 166872;
  Best Local Similarity 54.9%; Pred. No. 6.9;
  Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;
  Oy 43 tctcgattcgctcttttcaatacttcgggaaatagatgtgaaacccctataaac 102
  Db 144050 tcttataatttgactcttatttaattcttcagaaataaactgacacattatttaca 143991
  Oy 103 gcgggttttcgaagaacatgcctagatcatgtatgacacacatgactaagcaaaagt 162

```

Db 143990 GTTTTGGAAATTAAAGCTTAGAGACATTTATTAAGCTAAAGCTAAAGTAAATTAATT 143931

QY 163 gc 164

Db 143930 TC 143929

RESULT 15

AC010636

AC010636 179393 bp DNA

PRI 31-MAY-2000

DEFINITION Homo sapiens chromosome 19 clone CTD-2332E11, complete sequence.

AC010636

AC010636.6 GI:8122233

KEYWORDS

HTG.

SOURCE

human.

ORGANISM

Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS

1 (bases 1 to 179393) DOE Joint Genome Institute and Stanford Human Genome Center.

JOURNAL

Direct Submission

REFERENCE

2 (bases 1 to 179393)

AUTHORS

DOE Joint Genome Institute.

JOURNAL

Direct Submission

REFERENCE

Submitted (16-SEP-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

AUTHORS

3 (bases 1 to 179393) DOE Joint Genome Institute and Stanford Human Genome Center.

JOURNAL

Direct Submission

REFERENCE

Submitted (31-MAY-2000) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

AUTHORS

On May 31, 2000 this sequence version replaced gi:7711484.

JOURNAL

Draft Sequence Produced by DOE Joint Genome Institute

COMMENT

www.jgi.doe.gov

FINISHING

Completed at Stanford Human Genome Center

WWW

www.shgc.stanford.edu

QUALITY

Phrap Quality >=40 99.7% of Sequence;

ESTIMATED

Total Number of Errors is 1.2.

LOCATION

1. 179393

FEATURES

Location/Qualifiers

SOURCE

1. 179393

ORGANISM

"Homo sapiens"

DB_XREF

"taxon:9606"

CHROMOSOME

"19"

CLONE

"CTD-2332E11"

BASE COUNT

52588 a 36507 c 35542 g 54756 t

ORIGIN

Query Match 17.7%; Score 34; DB 9; Length 179393;

Best Local Similarity 54.9%; Pred. No. 7;

Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 43 tgcgcgattcgtcttcaaccctcggggaatagatgtgaaacctataaac 102

Db 128506 TGTAAATTTACCTATTAAATCTTCAGAAAAATTAACATATTATTACAA 128565

QY 103 gcgggtttcgcgaacatgcgtatcatatgacacacatgagcagaagt 162

Db 128566 GTTTTGGAAATTAAAGCTTAGAGACATTTATTAAGCTAAAGTAAATTAATT 128625

QY 163 gc 164

Db 128626 TC 128627

Search completed: March 28, 2002, 15:16:11
Job time: 7935 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2002, 13:16:56 ; Search time 240.01 Seconds

(without alignments)
685,832 Million cell updates/sec

Title: US-09-816-391A-1_COPY_1_192

Sequence: 1 gctggcgcgcgcgccatga.....tgaccacaagaagatgcttt 192

Scoring table: IDENTITY-NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 segs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08

Maximum Match 100%
Listing first 45 summaries

Database :

N.Geneseq.1101.*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	33.4	17.4	50000	21	AAH64140
2	32.6	17.0	278	21	AAC27526
3	30.4	15.8	1569	22	AAH52605
4	30.4	15.8	3204	22	AAH54121
5	29.2	15.2	3077	21	AAH46670
6	29.2	15.2	3092	14	AAO48934
7	29.2	15.2	3126	16	AAO90112
8	29.2	15.2	3202	21	AAC78077
9	29.2	15.2	3253	14	AAO48935
10	29.2	15.2	3324	22	AAH12765
11	29.2	15.2	3405	22	AAH89937

12	29.2	15.2	3448	22	AAH72618
13	29.2	15.2	160552	22	AAH02697
C 14	29	15.1	860	21	AAC74426
C 15	29	15.1	1444	22	AAH34227
C 16	29	15.1	1444	22	AAH05171
C 17	29	15.1	92407	22	AAH28549
C 18	28.6	14.9	16386	22	AAH63695
C 19	28	14.6	2882	22	AAH18508
C 20	27.8	14.5	730	22	AAH21896
C 21	27.8	14.5	1179	22	AAH52400
C 22	27.8	14.5	1257	15	AAO78141
C 23	27.8	14.5	1297	21	AAC48453
C 24	27.8	14.5	1311	21	AAC43114
C 25	27.8	14.5	1603	21	AAC33162
C 26	27.8	14.5	4434	22	AAH54373
C 27	27.8	14.5	5391	19	AAV71729
C 28	27.8	14.5	5407	19	AAV71738
C 29	27.8	14.5	24183	22	AAH21771
C 30	27.6	14.4	285	21	AAH78563
C 31	27.6	14.4	967	21	AAH45504
C 32	27.6	14.4	1623	21	AAH78542
C 33	27.4	14.3	484	22	AAH84040
C 34	27.4	14.3	936	22	AAH58252
C 35	27.4	14.3	936	22	AAH58254
C 36	27.4	14.3	936	22	AAH58257
C 37	27.4	14.3	936	22	AAH58259
C 38	27.4	14.3	936	22	AAH58262
C 39	27.4	14.3	936	22	AAH58255
C 40	27.4	14.3	1143	22	AAH67361
C 41	27.4	14.3	1266	22	AAH67831
C 42	27.4	14.3	1386	22	AAH07063
C 43	27.4	14.3	349980	22	AAH68531
C 44	27.4	14.3	580073	18	AAH58840
C 45	27.2	14.2	750	21	AAC47825

ALIGNMENTS

RESULT 1	AAH64140/C
ID	AAH64140 standard; DNA; 50000 BP.
XX	AC AAA64140;
XX	20-DEC-2000 (first entry)
DE	Nucleotide sequence of a beta-tubulin antigen.
XX	Beta-tubulin antigen; inner ear protein; Meniere's disease; autoantibody;
KW	chronic ear disease; autoimmune disease; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200050593-A1.
XX	
PD	31-AUG-2000.
XX	
PF	25-FEB-2000; 2000WO-US04795.
XX	
PR	25-FEB-1999; 99US-0121549.
XX	
PA	(UYTE-) UNIV TENNESSEE RES CORP.
XX	
PI	YOO TJ;
XX	
DR	WPI; 2000-558400/51.
XX	
PT	New beta-tubulin antigen in the membranous structure of the inner ear,
PT	reactive with antibodies of patients with Meniere's disease, for
PT	diagnosing Meniere's disease and distinguishing this disease from other
XX	autoimmune ear diseases

PS Claim 3; Page 74-97; 115pp; English.

XX The present sequence encodes a beta-tubulin antigen. The protein is

CC an antigen of the membranous structure of the inner ear protein, and

CC is reactive with antibodies from patients having Meniere's disease.

CC Meniere's disease is a chronic ear disease with unknown etiology.

CC Serum from patients suffering from this disease contain autoantibodies

CC against a 30 kDa cochlear protein antigen. The disease is believed to be

CC an autoimmune disease. The beta-tubulin antigen is useful as a target

CC substance in diagnosing or detecting Meniere's disease and in

CC distinguishing this disease from other autoimmune ear diseases.

XX

Sequence 50000 BP; 17281 A; 9480 C; 8791 G; 14448 T; 0 other;

SO

Query Match 17.4%; Score 33.4; DB 21; Length 50000;

Best Local Similarity 52.5%; Pred. No. 0.41;

Matches 73; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

OY 42 ttgtctgattcgtctatttcaatccttcggggaatagatgtgaaacccctataaaa 101

DB 26376 TTTTCTACCTTGCCTCTTACACAGCTACACCAATGATCTGTCACAGCAATGAG 26317

OY 102 cgcgggttttcgcaagaacatcgctagatcatgtgacacaatgagactagaacaaag 161

DB 26316 CCATTTTTCAGCATTAATTTTCCTTCCATGATTAATAATGAAATCAGCTAAGCATGAT 26257

OY 162 tgcctgtccctgaccccaa 180

DB 26256 TTCTTACTGCTGTGACCAA 26238

RESULT 2

AAC27526

ID AAC27526 standard; cDNA; 278 BP.

XX AAC27526;

AC 06-OCT-2000 (first entry)

XX

DT Human secreted protein 5' EST, SEQ ID NO: 31601.

XX

DE Human secreted protein 5' EST, SEQ ID NO: 31601.

XX

KM Human; 5' EST: expressed sequence tag; secreted protein; cDNA isolation;

XX gene therapy; chromosome mapping; ss.

XX

OS Homo sapiens.

XX

PN EP1033401-A2.

XX

PD 06-SEP-2000.

XX

PF 21-FEB-2000; 2000EP-0200610.

XX

PR 26-FEB-1999; 99US-0122487.

XX

PA (GEST) GENSET.

XX

PI Dumas Milne Edwards J, Duclert A, Giordano J;

XX

DR WPI; 2000-500381/45.

XX

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for

XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for

XX diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX

PS Claim 1; SEQ ID 31601; 71pp + CD-ROM; English.

XX

XX The present sequence is one of a large number of 5' ESTs derived from

CC mRNAs encoding secreted proteins. No ORF has yet been conclusively

CC identified within the present sequence. The 5' ESTs were prepared from

CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST

CC sequences usually correspond mainly to the 3' untranslated region (UTR)

CC of the mRNA because they are often obtained from oligo-dT primed cDNA

CC libraries. Such ESTs are not well suited for isolating cDNA sequences

CC derived from the 5' ends of mRNAs and even in those cases where longer

CC cDNA sequences have been obtained, the full 5' UTR is rarely included.

CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be

CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used

CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.

CC They are used to obtain upstream regulatory sequences and to design

CC expression and secretion vectors.

XX

Sequence 278 BP; 82 A; 44 C; 43 G; 108 T; 1 other;

SO

Query Match 17.0%; Score 32.6; DB 21; Length 278;

Best Local Similarity 51.0%; Pred. No. 0.08;

Matches 77; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

OY 18 ttaagtgcgttacaagaacatactgtctgattcgtctatttcaatccttcgggaa 77

DB 99 tcattggtattgtttcttcctttagtattaaatgctgtattcaataattkgygac 158

OY 78 atagatgtgaaacccctataaagcggttttcgagaacaatcgctagtatcatgt 137

DB 159 atatatgtcaattcaacatatatacagatgtgtatcaggtgacattccaaagtgcactg 218

OY 138 atgacaacatgactaagaacaaagtctgt 168

DB 219 ctgagttagaagatcatgtattgttaattc 249

RESULT 3

AAH52605

ID AAH52605 standard; DNA; 1569 BP.

XX AAH52605;

AC 03-SEP-2001 (first entry)

XX

DT S. epidermidis open reading frame nucleotide sequence SEQ ID NO:603.

XX

DE S. epidermidis open reading frame nucleotide sequence SEQ ID NO:603.

XX

KM Staphylococcus epidermidis SRI strain; infection; diagnosis;

XX vaccination; endocarditis; ds.

XX

OS Staphylococcus epidermidis.

XX

PN WO200134809-A2.

XX

PD 17-MAY-2001.

XX

PF 09-NOV-2000; 2000WO-US30782.

XX

PR 09-NOV-1999; 99US-0164258.

XX

PA (GLAX) GLAXO GROUP LTD.

XX

PI Kimmery MJ;

XX

XX WPI; 2001-316495/33.

XX

DR P-PSDB; AAG81755.

XX

PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,

XX useful for vaccinating against infections, e.g. endocarditis -

XX

PS Claim 8; Page 195; 2188pp; English.

XX

XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides

CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.

CC (I) and (II) can have antibacterial activity and therefore can be used

CC in vaccination. The nucleic acids (I) may be used to produce the

CC S. epidermidis polypeptides (II) via the production of vectors

CC containing them which are used to produce hosts cells which express the

CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be

CC used to vaccinate subjects and to raise antibodies against the bacteria.

CC The polypeptides may also be used to assay for other inhibitors of their

XX Sequence 3253 BP; 1106 A; 599 C; 665 G; 883 T; 0 other;
SQ

Query Match 15.2%; Score 29.2; DB 14; Length 3253;
Best Local Similarity 59.8%; Pred. No. 3.8;
Matches 49; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

OY 89 aaccctataaagcggttctgcagaacatgcgtatcatcatgatgacaacatg 148
|||||
DB 620 aaccattatcctaatgggtgtctacgttaattgtctgaatcattcattcgggaaccag 679

OY 149 gactaagcaaaagtcgtctcc 170
| |||||
DB 680 attgcacaacatggtgtctctcc 701

RESULT 10

AA012765
ID AAD12765 standard; CDNA; 3324 BP.

AC AAD12765;

DT 23-OCT-2001 (first entry)

DE Human OSF-2 (periostin) cDNA.

XX Human; OSF-2; periostin; antiinflammatory; nephrotrophic; renal disorder;

KW Immunoglobulin A nephropathy; IgAN; glomerulonephritis; inflammation;

KM mesangial proliferation; necrotising crescentic glomerulonephritis;

KW NCGN; minimal change disease; sclerosis; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS

FT 33..2543
/*tag= a
/product= "Human OSF-2 protein"

PN W0200157062-A1.

XX 09-AUG-2001.

XX 05-FEB-2001; 2001MO-US03654.

XX 03-FEB-2000; 2000US-0180136.

XX (GENE-) GENE LOGIC INC.

XX Munger WE, Sun H, Falk RJ;

XX WPI; 2001-488865/53.

XX P-PSDB; AA06520.

XX New polypeptide encoding OSF-2 (periostin); useful as a marker in renal

PT disease, e.g., immunoglobulin A nephropathy

XX Claim 1; Page 46-50; 57pp; English.

XX The patent discloses methods of correlating gene expression with
CC inflammatory diseases and renal disorders in healthy and disease
CC samples. The invention specifically relates to methods which monitor
CC the expression of OSF-2 (periostin). The OSF-2 proteins are useful
CC as markers in renal diseases such as immunoglobulin A nephropathy
CC (IgAN). Modulators of OSF-2 are used to treat IgAN, glomerulonephritis,
CC including mesangial proliferation; inflammation; necrotising crescentic
CC glomerulonephritis (NCGN); minimal change disease and sclerosis. The
CC present sequence is a CDNA encoding human OSF-2 (periostin) protein.

XX Sequence 3324 BP; 1146 A; 608 C; 676 G; 894 T; 0 other;

Query Match 15.2%; Score 29.2; DB 22; Length 3324;

Best Local Similarity 59.8%; Pred. No. 3.8;
Matches 49; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

OY 89 aaccctataaagcggttctgcagaacatgcgtatcatcatgatgacaacatg 148
|||||
DB 621 aaccattatcctaatgggtgtctacgttaattgtctgaatcattcattcgggaaccag 680

OY 149 gactaagcaaaagtcgtctcc 170
| |||||
DB 681 attgcacaacatggtgtctctcc 702

RESULT 11

AAH89937
ID AAH89937 standard; CDNA; 3405 BP.

AC AAH89937;

DT 01-OCT-2001 (first entry)

DE Human bone marrow CDNA, SEQ ID NO: 68.

XX Human; Bone marrow; antiinflammatory; cytostatic; neuroprotective;

KW antiviral; antibacterial; antifungal; anti-HIV; haemostatic;

KW immunosuppressive; gene therapy; cytokine cell proliferation;

KW cell differentiation modulator; immune disorder; infection; cancer;

KW human immunodeficiency virus; HIV; autoimmune disorder; haemophilia; ss.

XX Homo sapiens.

XX W0200153453-A2.

XX 26-JUL-2001.

XX 23-DEC-2000; 2000MO-US34960.

XX 21-JAN-2000; 2000US-0488725.

XX 25-APR-2000; 2000US-0552317.

XX 09-JUL-2000; 2000US-0598042.

XX 19-JUL-2000; 2000US-0620312.

XX 13-AUG-2000; 2000US-0653450.

XX 14-SEP-2000; 2000US-0662191.

XX 19-OCT-2000; 2000US-0693036.

XX 30-NOV-2000; 2000US-0250583.

XX (HYSE-) HYSEO INC.

XX Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Chen R, Ma Y;

XX Ren F, Wang J, Werhman T, Xu C, Xue AJ, Yang Y, Zhang J;

XX Zhao QA, Zhou P, Drmanac RT;

XX WPI; 2001-488707/53.

XX P-PSDB; AA00818.

XX Novel bone-marrow-expressed polynucleotides and polypeptides, useful

PT for treating e.g. cancer and immune deficiency disorders

XX Claim 1; Page 252-253; 648pp; English.

XX The present sequence is one of 251 novel human polynucleotides
CC expressed in the bone marrow. The polynucleotide and the
CC polypeptide encoded by it are useful in the treatment of various
CC immune deficiencies and disorders. The deficiencies and disorders may
CC be genetic, may be caused by a viral (e.g. HIV), bacterial or fungal
CC infection, or may result from an autoimmune disorder, a coagulation
CC disorder (e.g. haemophilia), inhibition of tumour cell proliferation,
CC suppression of an inflammatory response or treatment of a nervous
CC system disorder such as Alzheimer's disease. Detection of the presence
CC or increased expression of the polynucleotide or the protein it
CC encodes is useful for the diagnosis and/or prognosis of one
CC or more types of cancer. The polynucleotide and polypeptide can be
CC used as nutritional sources or supplements and in the screening of
CC chemical compounds as potential drugs.

CC disorders, cerebrovascular disorders, angiogenesis, nervous system
CC disorders, infections caused by bacteria, viruses and fungi and ocular
CC disorders. The polypeptides can also be used to aid wound-healing and,
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC maintain organs before transplantation, for supporting cell culture of
CC primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities, fat content, lipid, protein,
CC carbohydrate, vitamins, minerals, cofactors and other nutritional
CC components. AAC74387 to AAC74395 and AAB40150 represent sequences used in
CC the exemplification of the present invention.
XX
SQ Sequence 860 BP; 293 A; 161 C; 128 G; 278 T; 0 other;

Query Match 15.1%; Score 29; DB 21; Length 860;
Best Local Similarity 57.0%; Pred. No. 2.5;
Matches 53; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
Qy 76 aaatagatgtaaaacccctataaaacgcgggttcgagaacatgcgtagtatcat 135
Db 266 AATAATATAGGAAATCTCTGTTAAGCACCTGCTCCGTAAGTAATGAGAGAGATCAT 207
Qy 136 tgatgacaacatgactaagcaaaagtcttgt 168
Db 206 CATGCACTACATATATATAGAAATAAGTATTTT 174

RESULT 15

AAH34227/C

ID AAH34227 standard; CDNA; 1444 BP.

XX AAH34227;

XX 03-SEP-2001 (first entry)

XX Human colon cancer antigen encoding CDNA SEQ ID NO:1309.

XX Human colon cancer antigen encoding CDNA SEQ ID NO:1309.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;

XX colorectal carcinoma; ss.

XX Homo sapiens.

XX MO200122920-A2.

XX 05-APR-2001.

XX 28-SEP-2000; 2000MO-US26524.

XX 29-SEP-1999; 99US-0157137.

XX 03-NOV-1999; 99US-0163280.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Barash SC, Birse CE, Rosen CA;

XX WPI; 2001-235357/24.

XX P-PSDB; AAG74822.

XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides;

XX useful for preventing, diagnosing and/or treating colorectal cancers -

XX Claim 1; Page 3046; 9803pp; English.

CC AAH32943 to AAH37195 and AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing

CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated P,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ Sequence 1444 BP; 487 A; 202 C; 227 G; 527 T; 1 other;

Query Match 15.1%; Score 29; DB 22; Length 1444;
Best Local Similarity 52.0%; Pred. No. 3.1;
Matches 65; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
Qy 56 tatttcaataccttcggggaatagatgtaaaacccctataaaacgcgggtttcgca 115
Db 191 TATTACTGAAACCTTCCATCTACTACAGAGTCTGTTTAAATATTATTAGTTTACAA 132
Qy 116 gaacatgctgtagatcatcattgatgacacatggaactgaagaaagtctgtccctga 175
Db 131 GAATCATTTCTAATGAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 72
Qy 176 cccaa 180
Db 71 TCANA 67

Search completed: March 28, 2002, 15:15:31
Job time: 715 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2002, 13:05:56 ; Search time 112.49 Seconds
(without alignments)
386.557 Million cell updates/sec

Title: US-09-816-391A-1_COPY_1_192

Sequence: 192
1 gctggcgcgcgccatga.....tgaccacaagaagatgcttt 192

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 11323899 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_NA: *
1: /cgn2_6/prodata/2/lna/5A.COMB.seq: *
2: /cgn2_6/prodata/2/lna/5B.COMB.seq: *
3: /cgn2_6/prodata/2/lna/5A.COMB.seq: *
4: /cgn2_6/prodata/2/lna/5B.COMB.seq: *
5: /cgn2_6/prodata/2/lna/PCRTUS.COMB.seq: *
6: /cgn2_6/prodata/2/lna/Backfilssl.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29.2	15.2	3092	1	US-08-426-627-3
2	29.2	15.2	3126	2	US-08-426-627-3
3	29.2	15.2	3253	1	US-08-426-627-5
4	27.8	14.5	730	4	US-08-943-731-128
5	27.8	14.5	1257	1	US-08-943-731-128
6	27.8	14.5	24183	4	US-08-943-731-3
7	27.8	14.5	1009	6	5223394-8
8	27.8	14.1	1244	2	US-08-204-288-3
9	27.8	14.1	3182	1	US-08-188-582-12
10	27.8	14.1	3182	1	US-08-646-715-12
11	27.8	14.1	3852	1	US-08-306-546C-1
12	27.8	14.1	3852	2	US-08-530-524A-1
13	26.2	13.6	1610	4	US-09-347-803-19
14	26.2	13.6	2801	3	US-08-747-221B-30
15	26.2	13.6	2801	3	US-08-747-221B-32
16	26.2	13.6	2801	4	US-09-005-051-30
17	26.2	13.6	2801	4	US-09-005-051-32
18	26.2	13.6	2836	3	US-08-747-221B-24
19	26.2	13.6	2836	4	US-08-747-221B-26
20	26.2	13.6	2836	4	US-09-005-051-24
21	26.2	13.6	2836	4	US-09-005-051-26
22	25.8	13.4	885	2	US-08-222-719-7
23	25.8	13.4	885	2	US-08-470-925-7
24	25.8	13.4	885	2	US-08-471-613-7
25	25.8	13.4	885	2	US-08-471-613-7
26	25.6	13.3	735	3	US-09-193-191-3
27	25.6	13.3	1062	4	US-09-371-913A-5

28	25.6	13.3	2172	1	US-08-158-732-50	Sequence 50, Appl
29	25.6	13.3	2172	1	US-08-611-928-50	Sequence 50, Appl
30	25.6	13.3	2172	2	US-09-173-891-50	Sequence 50, Appl
31	25.6	13.3	3267	3	US-08-257-963B-12	Sequence 12, Appl
32	25.6	13.3	3267	5	US-08-257-963B-12	Sequence 12, Appl
33	25.6	13.3	3267	5	US-08-947-823-2	Sequence 2, Appl
34	25.6	13.3	22481	5	US-08-947-823-2	Sequence 2, Appl
35	25.6	13.3	51952	3	US-08-947-823-1	Sequence 43, Appl
36	25.4	13.2	1200	3	US-08-854-531-5	Sequence 1, Appl
37	25.4	13.2	1200	5	US-08-854-531-5	Sequence 5, Appl
38	25.2	13.1	420	1	US-08-680-726A-65	Sequence 5, Appl
39	25.2	13.1	420	3	US-09-092-409-65	Sequence 65, Appl
40	25.2	13.1	2044	1	US-08-680-726A-63	Sequence 63, Appl
41	25.2	13.1	2044	1	US-08-680-726A-64	Sequence 63, Appl
42	25.2	13.1	2044	3	US-09-092-409-63	Sequence 63, Appl
43	25.2	13.1	2044	3	US-09-092-409-64	Sequence 64, Appl
44	25.2	13.1	3758	3	US-08-323-477-1	Sequence 1, Appl
45	25.2	13.1	8532	1	US-08-452-655B-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-426-627-3
; Sequence 3, Application US/08426627
; Patent No. 5756664
; GENERAL INFORMATION:
; APPLICANT: Amano, Egon
; APPLICANT: Okawara-Hamamoto, Yoko
; APPLICANT: Kikuno, Reiko
; APPLICANT: Takeshita, Sunao
; APPLICANT: Tezuka, Kenichi
; TITLE OF INVENTION: No. 5756664e1 Protein with Bone Formation
; TITLE OF INVENTION: Ability and Process for Its Production.
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESS: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/426,627
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/036,841
; FILING DATE: 25-MAR-1993
; APPLICATION NUMBER: JP 4-71501
; FILING DATE: 27-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Hammond, Alan W.
; REGISTRATION NUMBER: 35,178
; REFERENCE/DOCKET NUMBER: 02481-1285-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3092 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORIGINAL SOURCE:

ORGANISM: Homo sapiens
TISSUE TYPE: Placenta
FEATURE:
NAME/KEY: CDS
LOCATION: join(38..2375)
FEATURE:
NAME/KEY: mat_peptide
LOCATION: join(101..2375)
US-08-426-627-3

Query Match 15.2%; Score 29.2; DB 1; Length 3092;
Best Local Similarity 59.8%; Pred. No. 1;
Matches 49; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 89 aaccctataaagcggttctgcagaaacatgcgctagatcattgagacaatg 148
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 626 AACCATATATCCTAATGCGGTTCACGTTAATTGCTCGAATCATGCAGAACACG 685
QY 149 gactaagcaaaagtgtgttc 170
| ||||| | |||||
Db 686 ATTGCAACAATAATGCTGTGTCC 707

RESULT 2
US-08-477-396A-3

Sequence 3, Application US/08477396A
Patent No. 5872235
GENERAL INFORMATION:
APPLICANT: Chen, Ian Bo
APPLICANT: Bao, Shideng
APPLICANT: Liu, Yuan
TITLE OF INVENTION: A NOVEL TUMOR MARKER AND NOVEL METHOD OF
NUMBER OF SEQUENCES: 19
TITLE OF INVENTION: ISOLATING SAME
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Weingarten, Schurgin, Gagneblin & Hayes
STREET: Ten Post Office Square
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,396A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/146,488
FILING DATE: 29-OCT-1993
APPLICATION NUMBER: US 08/448,388
FILING DATE: 28-MAY-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/12502
FILING DATE: 31-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: Heine, Holliday C.
REGISTRATION NUMBER: 34,346
REFERENCE/DOCKET NUMBER: DFCI-333BX
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-2290
TELEFAX: (617) 451-0313
INFORMATION FOR SEQ. ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3126 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA

HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 43..2376
US-08-477-396A-3

Query Match 15.2%; Score 29.2; DB 2; Length 3126;
Best Local Similarity 59.8%; Pred. No. 1;
Matches 49; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 89 aaccctataaagcggttctgcagaaacatgcgctagatcattgagacaatg 148
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 649 AACCATATATCCTAATGCGGTTCACGTTAATTGCTCGAATCATGCAGAACACG 708
QY 149 gactaagcaaaagtgtgttc 170
| ||||| | |||||
Db 709 ATTGCAACAATAATGCTGTGTCC 730

RESULT 3
US-08-426-627-5

Sequence 5, Application US/08426627
Patent No. 5756664
GENERAL INFORMATION:
APPLICANT: Amann, Egon
APPLICANT: Otawara-Hamamoto, Yoko
APPLICANT: Kikuno, Reiko
APPLICANT: Takeshita, Sunao
APPLICANT: Tezuka, Kenichi
TITLE OF INVENTION: No. 5756664el Protein with Bone Formation
NUMBER OF SEQUENCES: 24
TITLE OF INVENTION: Ability and Process for Its Production.
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/426,627
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/036,841
FILING DATE: 25-MAR-1993
APPLICATION NUMBER: JP 4-71501
FILING DATE: 27-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Hammond, Alan W.
REGISTRATION NUMBER: 35,178
REFERENCE/DOCKET NUMBER: 02481-1285-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ. ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 3253 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
CELL TYPE: osteosarcoma

```

:      FEATURE:      CDS
:      NAME/KEY:      join(32...2540)
:      LOCATION:
:      FEATURE:
:      NAME/KEY:      mat_peptide
:      LOCATION:      join(97...2540)
:
US-08-426-627-5

```

Query Match	15.2%	Score 29.2;	DB 1;	Length 3253;
Best Local Similarity	59.8%;	Pred. No. 1.1;		
Matches 49;	Conservative 0;	Mismatches 33;	Indels 0;	Gaps 0;

Oy	89	aacccttataaacgcgggttttcgcacaacatgycgtatgatcatgatgaacaatg	149
Db	620	AACCATTTCTTAATGGGCTGTCACTTAAATTGTGTCGAATCAGCCATCGGGAACCAG	675
Oy	149	gactcaagcaaaagtgccttgctc	170
Db	680	ATTGCACAAAATGGTGTTGCC	701

RESULT 4
US-08-943-731-128/c
Sequence 128 Application US/08943731
Patent No. 6265157
GENERAL INFORMATION:
APPLICANT: PROCKOP, DARWIN J.
APPLICANT: SPOTILA, LORETTA D.
APPLICANT: DELTAS, CONSTANTINOS D.
APPLICANT: SEREDA, LARISA
APPLICANT: LARSON, ANDREA W.
APPLICANT: PACK, MICHAEL
APPLICANT: COLIGE, ALAIN
APPLICANT: EARLY, JAMES
APPLICANT: KORRKO, JARMO
APPLICANT: ALA-KOKKO, LEENA, et al.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING
TITLE OF INVENTION: ALTERED TYPE I OR TYPE IX COLLAGEN GENE SEQUENCES
NUMBER OF SEQUENCES: 666
CORRESPONDENCE ADDRESS:
ADDRESSEE: PANTICH SCHWARZE JACOBS & NADEL, P. C.
STREET: ONE COMMERCE SQUARE, 2005 MARKET STREET, 22ND
STREET: FLR.
CITY: PHILADELPHIA
STATE: PA
COUNTRY: USA
ZIP: 19103-7086
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/943,731
FILING DATE: 03-OCT-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/212,322
FILING DATE: 14-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/803,628
FILING DATE: 03-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: DOYLE LEARY Ph. D., KATHRYN
REGISTRATION NUMBER: 36,317
REFERENCE/DOCKET NUMBER: 9596-27
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-965-1284
TELEFAX: 215-567-2991
TELEX: 831-494
INFORMATION FOR SEQ ID NO: 128:
SEQUENCE CHARACTERISTICS:

```

; LENGTH: 730 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-943-731-128

```

Query Match	14.5%	Score 27.8;	DB 4;	Length 730;
Best Local Similarity	65.1%;	Pred. No. 1.8;		
Matches 41;	Conservative 0;	Mismatches 22;	Indels 0;	Gaps 0;

Oy	130	tatcatgatgacaacacttgagcttaagcaaaaagtcttgcgccctcgaccgaaccaagaagatagc	189
Db	374	TTTCAGAAGTGACAGCCCTTGACCAAAGCTCAGCGTGGTATGCCCTGCCTCTGTGAAGGCTGC	315
Oy	190	ttt	192
Dd	314	TCT	312

```

RESULT 5
US-08-330-154-1
: Sequence 1, Application US/08330154
: Patent No. 5587307
: GENERAL INFORMATION:
: APPLICANT: Alborn Jr., William E
: APPLICANT: Hoskins, Joann
: APPLICANT: Skatrud, Paul L
: APPLICANT: Unal, Serhat
: TITLE OF INVENTION: FEWA GENE OF STAPHYLOCOCCUS EPIDERMIDIS,
: TITLE OF INVENTION: FEWA PROTEIN, AND VECTORS AND MICROORGANISMS COMPRISING
: TITLE OF INVENTION: THE FEWA GENE
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Patent Division/AEH
: STREET: Lilly Corporate Center
: CITY: Indianapolis
: STATE: Indiana
: COUNTRY: USA
: ZIP: 46285
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/330,154
: FILING DATE: 27-OCT-1994
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/208,925
: FILING DATE: 09-MAR-1994
: APPLICATION NUMBER: US/08/057,163
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Hamilton, Amy E
: REGISTRATION NUMBER: 33,894
: REFERENCE/DOCKET NUMBER: X-8894
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 317-276-3169
: TELEFAX: 317-276-1294
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1257 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..1257
:
US-08-330-154-1

```

RESULT 8
US-08-204-288-3/C
Sequence 3, Application US/08204288
Patent No. 5959178
GENERAL INFORMATION:
APPLICANT: VAN DOORSSELAERE, Jan
APPLICANT: FRITIG, Bernard J.M.
APPLICANT: INZE, Dirk G.
APPLICANT: JOUANIN, Lise
APPLICANT: KNIGHT, Mary E.
APPLICANT: VAN MONTAGU, Marc
APPLICANT: LEGRAND, Michel
TITLE OF INVENTION: MODIFICATION OF LIGNIN SYNTHESIS IN
TITLE OF INVENTION: PLANTS
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
STREET: 1100 New York Avenue, N.W.
CITY: Washington
STATE: D. C.
COUNTRY: U.S.A.
ZIP: 20005-3518
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/204,288
FILING DATE: 10-MAR-1994

[illegible]

Patent No. 6274379

; Patent No. 6274379
; GENERAL INFORMATION:

;	NAME/KEY:	CDS
;	LOCATION:	99..1886
US-08-747-221B-30		

QY 120 catgctgctatcaattgacacatgacaaagtcttgc 170
 | | | | | | | | | | | | | | | | | |
 Db 697 GAAGAAAAGCTTCATTATGAGGCATCTCAGTCATBAAAAAGTCTTTTCC 747

Search completed: March 28, 2002, 15:10:40
Job time: 7484 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2002, 12:36:06 ; Search time 2300.01 Seconds
(without alignments)
897.035 Million cell updates/sec

Title: US-09-816-391A-1_COPY_1_192

Perfect score: 192

Sequence: 1 gctggggcgcgcgcgcacatga.....tgaccagaagatgcttt 192

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estlin:*
4: em_estlom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: gb_est1:*
11: gb_est2:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rtd:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	36.4	19.0	749	13	BH039236
C 2	34.6	18.0	852	11	BG542199
C 3	34.2	17.8	600	13	AZ964976
C 4	33.8	17.6	945	13	CNS079HE
C 5	33.6	17.5	845	13	CNS0757U
C 6	33.2	17.3	611	13	AO017410
C 7	33.2	17.2	630	13	AQ268787
C 8	32.8	17.1	523	13	AZ931427
C 9	32.6	17.0	957	13	CNS04NR7
C 10	32.6	17.0	552	13	BG062885
C 11	32.6	17.0	958	13	CNS0480J
C 12	32.6	17.0	1101	13	CNS007P

13	32.6	17.0	2062	12	AK005440
14	32.6	17.0	2366	12	AK004888
15	32.4	16.9	510	13	AO787593
16	32.2	16.8	542	13	AO564194
C 17	32.2	16.8	746	13	AZ900491
C 18	32.2	16.7	439	13	AZ099361
C 19	31.6	16.5	291	10	BB485889
C 20	31.4	16.4	236	10	BB083723
C 21	31.4	16.4	1511	11	BF234077
C 22	31.2	16.2	434	11	BF547099
C 23	31.2	16.2	650	10	AV733876
C 24	31.2	16.2	677	13	AO196619
C 25	31.2	16.2	973	13	CNS02ETM
C 26	31.1	16.1	621	10	AV733687
C 27	31.1	16.1	680	13	AO927062
C 28	30.8	16.0	769	11	BG068661
C 29	30.6	15.9	318	10	AV296697
C 30	30.6	15.9	320	11	BI202015
C 31	30.6	15.9	354	10	AW443473
C 32	30.6	15.9	613	10	AI779343
C 33	30.6	15.9	615	10	AI779344
C 34	30.6	15.9	1101	13	CNS00BLA
C 35	30.4	15.8	408	10	BES81870
C 36	30.4	15.8	444	13	AO153210
C 37	30.4	15.8	445	10	BES79806
C 38	30.4	15.8	445	11	BG227023
C 39	30.4	15.8	617	13	AO486794
C 40	30.4	15.8	735	13	AO384366
C 41	30.4	15.8	777	11	BG849734
C 42	30.4	15.8	865	11	BF246893
C 43	30.2	15.7	471	11	BG047104
C 44	30.2	15.7	532	11	W81666
C 45	30.2	15.7	565	13	AO534021

ALIGNMENTS

RESULT 1
LOCUS BH039236/c
DEFINITION RPCI-24-254N24.TV RPCI-24 Mus musculus genomic clone RPCI-24-254N24
' DNA sequence.
ACCESSION BH039236
VERSION BH039236.1 GI:14816964
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE
AUTHORS Zhao,S., Niernan,W., Malek,J., Shatsman,S., Akiret,B., Levins,M., Tsegaye,G., Geer,K., Krol,M., Shvatsldayn,A., Gebregeorgis,E., Russell,D., de Jong,P. and Fraser,C.M.
TITLE Mouse BAC End Sequences from Library RPCI-24
JOURNAL Unpublished (1999)
COMMENT Other-GSSs: RPCI-24-254N24.TJ

CONTACT: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208

Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC

library availability, please contact Pierre de Jong

(pdejong@tigr.org). Clones may be purchased from BACPAC

Resources (<http://www.chori.org/bacpac/orderingframe.htm>). BAC end

plate: 254 row: N column: 24
Seq primer: T7
Class: BAC ends.

FEATURES

Location/Qualifiers

```

source
1. 749
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-254N24"
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"
/note="Vector: pTARBACK1. Site_1: BamHI; Site_2: BamHI.
RPCI-24 Mouse BAC Library produced by Pletter de Jong. The
library was cloned in the pTARBACK1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."

BASE COUNT      216 a      170 c      165 g      198 t

ORIGIN

Query Match      19.0%; Score 36.4; DB 13; Length 749;
Best Local Similarity 54.5%; Pred. No. 0.39;
Matches 73; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

Oy 13 gggcattgaagtggcttgcacagcataatctgtctgtatgcgtacttcaactctcg 72
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 456 ggcatctctgggtggcttccatgcacaaatcactgcgttgatgatgtccagatcttgc 397
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Oy 73 gggaaatgagtgtgaaaccccttaaaacgggggttttcgagagaatagcgctagat 132
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 396 atgcatttggtgtgaaatccatgaatgcagatgacctctccattttgaagccatgctgttat 337
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Oy 133 catgatgacacaa 146
      | | | | | | |
Db 336 ctctgattgcacaaa 323

RESULT 2
BG542199      852 bp      mRNA      EST      03-APR-2001
LOCUS      602571637F1 NIH_MGC_77 Homo sapiens cdna clone IMAGE:4696041 5',
DEFINITION      mRNA sequence.
ACCESSION      BG542199
VERSION      BG542199.1 GI:13534432
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 852)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@remail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMN at:
http://image.lnl.gov
Plate: L10CM1522 row: j column: 10
High quality sequence stop: 365.
Location/Qualifiers
1. 852
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4696041"
/clone_lib="NIH_MGC_77"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: lung; Vector: pDMR-LIB (Clontech); Site_1:
SfiI (ggcgccctcgccgc); Site_2: SfiI (ggccattgccc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGGCCATTATGACC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCCGAGGCGCCGACATG-dT(30)BN-3' (where B = A,

```

C, or G and N = A, C, G, or T). Average insert size 1.9 kb (range 0.5-4.0 kb). 12/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH/MCC Library."

BASE COUNT	215 a	260 c	199 g	178 t
ORIGIN				

Query Match	18.0%	Score 34.6	DB 11	Length 852
Best Local Similarity	55.4%	Pos. No. 1.5		
Matches 67	Conservative 0	Mismatches 54	Indels 0	Gaps 0
Oy	38 aatctgtcgtatcgtcgtatcttcaacaccttcggggaatagatgtaaacacctat	97		
Db	446 ACTATGCTCATATGCGCCACACGCACATFACATGGGAGCCTTAAGTGCACGCCCATAT	505		
Oy	98 aaacgcgggttttcgcgaagaacatgcgctatgatacttgatgacaacatggaacga	157		
Db	506 CAAGCGCGGCGCTATCCTATATCGGACAGTCACATTTTTCATACATGAGACTACCC	565		
Oy	158 a 158			
Db	566 A 566			

RESULT 3	AZ964976/c	LOCUS	600 bp	DNA	GSS	27-APR-2001
DEFINITION	2M0234D2R Mouse 10kb plasmid UNGC2M library					Mus musculus genomic
ACCESSION	AZ964976					
VERSION	AZ964976.1					GI:13836203
KEYWORDS	GSS.					
SOURCE	house mouse.					
ORGANISM	Mus musculus					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.					
AUTHORS	1 (bases 1 to 600)					
	Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.					
TITLE	Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts					
JOURNAL	Unpublished (2000)					
COMMENT	Contact: Robert B. Weiss University of Utah Genome Center Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA Tel: 801 585 5606 Fax: 801 585 7177 Email: ddunn@genetics.utah.edu Insert length: 10000 Std Error: 0.00 plate: 0234 row: D column: 24 Seq primer: CACACAGGAACACGCTATGACC Class: plasmid ends High quality sequence stop: 600.					
FEATURES	Location/Qualifiers					
SOURCE	1. 600					
	/organism="Mus musculus"					
	/strain="C57BL/6J"					
	/db_xref="taxon:10090"					
	/clone="UNG2M0234D24"					
	/clone_lib="Mouse 10kb plasmid UNGC2M library"					
	/sex="Female"					
	/lab_host="E. coli strain XL10-Gold, Tl-resistant, F-"					
	/note="Vector: PWD43nv; Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnaes/). The DNA was hydrodynamically sheared by repeated passage through a					

TITLE Direct Submission
JOURNAL Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Creteiloux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This GSS is part of a random genomic sequencing program of thirteen yeast species: *Saccharomyces bayanus* var. *uvaturum*, *Saccharomyces*

RESULT	10
LOCUS	BG062885
DEFINITION	BG062885 552 bp mRNA EST 25-JAN-2001
ACCESSION	L0959C09.5 NIA Mouse Newborn Kidney cDNA Library2 (Short) Mus
VERSION	musculus cdna clone L0959C09 5' mRNA sequence.
	BG062885
	BG062885.1 GI:12533811

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2002, 15:16:11 ; Search time 1551.88 Seconds
(without alignments)
1371.328 Million cell updates/sec

Title: US-09-816-391A-1_COPY_472_600

Perfect score: 129
Sequence: 1 tgcattctgctgcgtacgca.....tattccgagactagttcagcg 129

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_scs:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: em_hgo_hum:*
31: em_hgo_inv:*
32: em_hgo_rtd:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rtd:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	31	24.0	5832	1 AF105113	AF105113 Streptoco
2	31	24.0	18754	1 AF094575	AF094575 Streptoco
3	29.6	22.9	174118	2 AC024405	AC024405 Homo sapl
4	29.6	22.9	185341	2 AC069403	AC069403 Homo sapl
5	29.6	22.9	331109	2 AL162413	AL162413 Homo sapl
6	29.4	22.8	179759	2 AC092021	AC092021 Homo sapl
7	29.4	22.8	190919	2 AC091731	AC091731 Homo sapl
8	29.4	22.8	202224	2 AL356219	AL356219 Homo sapl
9	29.4	22.8	203674	2 AC025669	AC025669 Mus muscu
10	29.4	22.8	230838	2 AL591946	AL591946 Mus muscu
11	29.2	22.6	1877	10 AF176007	AF176007 Mus muscu
12	29.2	22.6	1928	10 AF192558	AF192558 Mus muscu
13	29.2	22.6	2193	1 RLU44387	RLU44387 Rhizobium 1
14	29.2	22.6	5263	1 RLRES	X98117 Rhizobium 1
15	29.2	22.6	21984	10 AB040292	AB040292 Mus muscu
16	29	22.5	138041	9 CNS01D01	AL133245 BAC seque
17	29	22.5	166318	9 CNS01D0G	AL133243 BAC seque
18	29	22.5	189008	2 AP001491	AP001491 Homo sapl
19	29	22.5	201657	2 AL596187	AL596187 Mus muscu
20	29	22.5	232190	2 AL596025	AL596025 Mus muscu
21	28.8	22.3	35694	8 SP8C23G7	AL035065 S.pombe c
22	28.8	22.3	152148	2 AC063952	AC063952 Homo sapl
23	28.8	22.3	185401	2 AC024927	AC024927 Homo sapl
24	28.6	22.2	738	1 BOV279043	AJ279043 Bacteroid
25	28.6	22.2	10353	1 AE004965	AE004965 Pseudomon
26	28.6	22.2	110000	2 LMFLCHR16_02	Continuation (3 of
27	28.6	22.2	171490	2 AL356092	AL356092 Homo sapl
28	28.6	22.2	180292	9 AL359269	AL359269 Human DNA
29	28.6	22.2	201320	2 AC073816	AC073816 Mus muscu
30	28.6	22.2	247196	2 AC073822	AC073822 Mus muscu
31	28.4	22.0	731	8 ZMHMGD1	Y08807 Z.mays mRNA
32	28.4	22.0	13104	1 AE000985	AE000985 Archaeogl
33	28.4	22.0	65762	2 AC079913	AC079913 Homo sapl
34	28.2	21.9	173515	2 AC073394	AC073394 Homo sapl
35	28.2	21.9	177642	2 HSDJ90810	AL122034 Homo sapl
36	28.2	21.7	332635	1 AP003005	AP003005 Mesophilic
37	28	21.7	2384	1 SRMSIK	Y08921 S.reticuli
38	28	21.7	143515	8 AP002526	AP002526 Oryza sat
39	27.8	21.6	110019	9 AC068860	AC068860 Homo sapl
40	27.8	21.6	145428	2 AC026068	AC026068 Homo sapl
41	27.8	21.6	147484	2 AL590088	AL590088 Homo sapl
42	27.8	21.6	203668	2 AL592148	AL592148 Homo sapl
43	27.6	21.4	2150	1 AF017790	AF017790 Homo sapl
44	27.6	21.4	8280	1 PSEPLAS	X66604 Pseudomonas
45	27.6	21.4	9617	1 RLY12758	LY12758 R.leguminos

ALIGNMENTS

RESULT 1
LOCUS AF105113 5832 bp DNA BCT 09-SEP-1999
DEFINITION Streptococcus pneumoniae type 19A putative oligosaccharide repeat unit transporter (cps19A) gene, partial cds; UDP-N-acetyl glucosamine-2-epimerase (cps19AK), glucose-1-phosphate thymidyl transferase (cps19AL), dTDP-4-keto-6-deoxyglucose-3,5-epimerase (cps19AM), dTDP-glucose-4,6-dehydratase (cps19AN), and dTDP-L-rhamnose synthase (cps19AO) genes, complete cds; and AL1A (alia) gene, partial cds.
ACCESSION AF105113
VERSION AF105113.1 GI:4406246
KEYWORDS

SOURCE Streptococcus pneumoniae.
ORGANISM Streptococcus pneumoniae
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae; Streptococcus.
REFERENCE 1 (bases 1 to 5832)
AUTHORS Morona,J.K., Morona,R. and Paton,J.C.
TITLE Comparative genetics of capsular polysaccharide biosynthesis in

JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL

Streptococcus pneumoniae types belonging to serogroup 19
J. Bacteriol. 181 (17), 5335-5364 (1999)
2 (bases 1 to 5832)
Morona, J.R., Morona, R. and Paton, J.C.
Submitted (10-NOV-1998) Molecular Microbiology Unit, Women's and
Children's Hospital, 72 King William Rd, Adelaide, SA 5006,
Australia

FEATURES
Source

gene
CDS

Location/Qualifiers
1..5832
/organism="Streptococcus pneumoniae"
/strain="19A2"
/serotype="19A"
/db_xref="taxon:1313"
/note="serogroup: 19"
1..818
/gene="cps19A"
/gene="cps19A"
/note="Cps19A"
/codon_start=3
/transl_table=11
/product="putative oligosaccharide repeat unit transporter"
/protein_id="AAD19911.1"
/db_xref="GI:4406247"
/translation="SKFIFNIKESFDNIRVGINLMSINASSLIIGIYRLGIONMNI
ETFGKVSLLTSLVSNLMTFINAGLVPLRPRKAEKLPKITSNLRVNLINFAIL
LIYPLKIVIDLMLPAYODALLFMPLIPPMVSVEGKALVINTYIKALRMRDILRVN
ALVNLISMGVITLTYTLINSLLELYIVTLALRSITAEILTSKLDVSKKDIIVLE
FLTLVPLISSWYPLIGLAVIVTIAAGLYILYLNHEDIKTYLAFKASKTSN"
883..1971
/gene="cps19A"
883..1971
/gene="cps19A"
/note="Cps19A"
/codon_start=1
/transl_table=11
/product="UDP-N-acetyl glucosamine-2-epimerase"
/protein_id="AAD19912.1"
/db_xref="GI:4406248"

/translation="MKKIMLVETGPRPAIKKCPVLNLEKKHEDEKMETVTCQHKEMV
SPVLDLGVVPDIDLETKANQTLSTTSTLEKIKVLEKQPDIVLHDDTTTAA
AALAAFLYGLKVGVEAGLRTYNQSPPEPFRQSTSIATYHFAPELAKENILE
GRENVYTGNTVIDALTTVQEDYTHLIDLANNRLLILTAHRENIGEMRMPRA
VKRYLNEYEDVKVYPIHKNPLVRETAETIGDTERIOIIEPLDLDHNNMNSYMT
LDSGGVQOEAPSLGKPVLMVROTTERPEGVAAGTLLKLVGIDEETIYQNRLLDDSG
EYKMSQASNPYNGDASOOIVQILKGI"
2134..3003
/gene="cps19A"
2134..3003
/gene="cps19A"
/note="Cps19A"
/codon_start=1
/transl_table=11
/product="glucose-1-phosphate thymidyl transferase"
/protein_id="AAD19913.1"
/db_xref="GI:4406249"

/translation="MKGIILAGSGCTLYLPLTRAASKOIMPYVDPKMIYPLSTMLA
GIDDLISYPPOLPRKDLIDDSSEGIKISTAEOPSPGLAOFLLIGEFIDDSV
ALILGDIYHPCGLSTMLQKAAKKEGATVGVQKQPERGVVETFDNNALSTIEK
PEYPRSNVAVTGLFYDNDVVEIAKOIKPSARGLEITDVNKAVLNGDLSVEIIMGSG
FAMLDGTGHTSLLEASQYIEYVOMVQVANLEISYRMGYSIREVLELAQPLKKN
EYGRYLLRIGEA"
3004..3597
/gene="cps19A"
3004..3597
/gene="cps19A"
/note="Cps19A"
/codon_start=1
/transl_table=11
/product="dUDP-4-keto-6-deoxyglucose-3,5-epimerase"

gene
CDS

BASE COUNT 1756 a 1083 c 1237 g 1756 t
ORIGIN

gene
CDS

Query Match 24.0%; Score 31; DB 1; Length 5832;
Best Local Similarity 59.8%; Pred. No. 17;
Matches 52; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

/protein_id="AAD19914.1"
/db_xref="GI:4406250"
/translation="MSDNFGKTLVVRKIDAPICLLERPDIVPHGDNRCGFKENOKK
MLPLGFPSPFAACKLQNNVSPSKRNVLRGIAHEWQKYSVADGKVLSSVVDLREG
ETFGNVQTEIDASKGIFVPRGVANGFVLSDTVSYLVNDYALTEKPRYARVNA
DPLAGLIEWENLEAEESEADKHPLLRDVKPLTDEL"
3609..4655
/gene="cps19A"
3609..4655
/gene="cps19A"
/note="Cps19A"
/codon_start=1
/transl_table=11
/product="GDP-glucose-4,6-dehydratase"
/protein_id="AAD19915.1"
/db_xref="GI:4406251"

/translation="MTEKILITVGAGFISNFWHYVYNNPDPVHTVLDKLYAGN
RANLEILGDRVELVGDIDADALVYKLAEMADIVHVAASHNDNSINDSPFIHTN
FIGYITLLEARKYDIRFHVSTDEYVGDLELRDLPGHGLGEKFAETKYPSP
YSSTKAASDLIVKAMVRSFGVKATISGNSNNGYOHIEKTIPTOINIGIKPKLY
GEGKNVRWIMHTNDHSSGVWTLTKGIGETYLIGADGKKNKVELLEIKEMQPD
AYDHVTRAGDLEKXAYDASKLDELGNKRPETTFPEAGLKETIKWYTDNDMMKSEK
AVERNYAKTQOYIR"
4659..4704
complement(4717..5571)
/gene="cps19A"
complement(4717..5571)
/gene="cps19A"
/note="Cps19A"
/codon_start=1
/transl_table=11
/product="GDP-L-rhamnose synthase"
/protein_id="AAD19916.1"
/db_xref="GI:4406252"

/translation="MILITGANGQLGTELRYLLNERVDYVANDVTEMDITNSEVAK
VFEAKPPLVYHCAATVADVADEGRLELFAINVTGTENYAKGSEKYGAILVYISTD
YVEDGKRPVQGEVEMVDIDPDOTEYGRTRKMGELVNLNLSOYH1IRTAWFGVYGRN
FVFTMOISLAETHTKTLIVYSDOYGRPTWTRTAEEMTVTENOKEYGYHHLSNDAAET
TWYDFASEILKESDVELLPVDSKFPKAKRPLNSTMSLAKAKATGVITPTWDLAKE
FYQOEVRK"
5817..5832
/gene="a11a"
5817..5832
/gene="a11a"
/codon_start=1
/transl_table=11
/product="A11a"
/protein_id="AAD19917.1"
/db_xref="GI:4406253"
/translation="MKSKR"

BASE COUNT 1756 a 1083 c 1237 g 1756 t
ORIGIN

Query Match 24.0%; Score 31; DB 1; Length 5832;
Best Local Similarity 59.8%; Pred. No. 17;
Matches 52; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

Db 2722 ATTACACAGCTTACCAAGGCTTACCTTAATCGTGTGACCTTCTGTGACCTGATGGG 2781

Db 2782 CGTGTGTTTGGCTGTGATACGGGA 2808

RESULT 2
AF094575 18754 bp DNA BCT 09-SEP-1999
LOCUS
DEFINITION
Streptococcus pneumoniae serotype 19A Dextr (dexA) gene, partial
sequence; capsular polysaccharide biosynthesis operon, complete
sequence; and oligopeptide permease A11a (a11a) gene, partial cds.

ACCESSION	AF094575
VERSION	AF094575.1
KEYWORDS	GI:3907597
SOURCE	Streptococcus pneumoniae
ORGANISM	Streptococcus pneumoniae
REFERENCE	Bacteria: Firmicutes; Bacillus/Clostridium group; Streptococcaceae; Streptococcus.
AUTHORS	1 (bases 1 to 18754)
TITLE	Morona, J. K., Morona, R. and Paton, J. C.
JOURNAL	Analysis of the 5' portion of the type 19a capsular locus identifies two classes of cpsC, cpsD, and cpsE genes in Streptococcus pneumoniae
MEDLINE	J. Bacteriol. 181 (11), 3599-3605 (1999)
PUBMED	99287847
REFERENCE	2 (bases 1 to 18754)
AUTHORS	Morona, J. K., Morona, R. and Paton, J. C.
TITLE	Comparative genetics of capsular polysaccharide biosynthesis in Streptococcus pneumoniae types belonging to serogroup 19
JOURNAL	J. Bacteriol. 181 (17), 5355-5364 (1999)
MEDLINE	99395043
REFERENCE	3 (bases 1 to 18754)
AUTHORS	Morona, J. K., Morona, R. and Paton, J. C.
TITLE	Submitted (23-SEP-1998) Microbial Microbiology Unit, Women's and Children's Hospital, King William Rd, North Adelaide, SA 5006, Australia
FEATURES	Location/Qualifiers
SOURCE	1. 18754
gene	/organism="Streptococcus pneumoniae"
CDS	/serotype="19a"
gene	/db_xref="taxon:1313"
CDS	<1..9
gene	/gene="dexB"
CDS	<1..9
gene	/gene="dexB"
CDS	/codon_start=1
gene	/transl_table=11
CDS	/product="DexB"
gene	/protein_id="AAC78662.1"
CDS	/db_xref="GI:3907598"
gene	/translation="VE"
CDS	complement(161..1647)
gene	/note="Insertion sequence; 3' truncated"
CDS	/rpl_family="IS1202"
gene	/rpl_type="dispersed"
CDS	1733..3178
gene	/gene="cps19a"
CDS	1733..16682
gene	/gene="capsular polysaccharide biosynthesis operon"
CDS	1733..3178
gene	/gene="cps19a"
CDS	/codon_start=1
gene	/transl_table=11
CDS	/product="putative regulatory protein Cps19a"
gene	/protein_id="AAC78663.1"
CDS	/db_xref="GI:3907599"
gene	/translation="MSRRFKRSQKRVSVNIIYLLVCFLLFLFTYNIILAF RYLNLYVPLVLYLVALLYKKAETFLVPSILVSSYSLFAVOOFGILNR LNNTSNSEYSISVAIVLADSEIENVQMLVATPTGDNENIOKLADIKSNDTIT VNOSSSLAYKXSIAGETKATVINSFENIIESEHPDYASKIKITTKPTKVEAP KISKDSFNITVSGIDTGPITSSRSYVNIILMTVNDTKKILITTPRQAVPIADG GNNOKKLTAGIYGVDSIHTLENIYVDINTYVRLNFTSFLKIDLLGCVVYNDG DFTSLNGKFFHPYGNVHLDEQALGFYRERYSLADGDRNQOKVIAIILKLTSS EALKNSRIIDSLDSDIOTNMPLETMINLVNAOESGCTYKVNQSDLKGRGRDLPY AMPDSNLVMEINDSSLASVKTAIDVLEGR"
gene	3180..3911
CDS	/gene="cps19a"
gene	/gene="cps19a"
CDS	3180..3911
gene	/gene="cps19a"
CDS	/codon_start=1
gene	/transl_table=11
CDS	/product="Cps19a"
gene	/protein_id="AAC78664.1"
CDS	/db_xref="GI:3907600"
gene	/translation="MIDHSHIVFDVDDGPKRSRESKALLAESYRQGVRTVTSNHR KQMEPEEKIAENFLOVREIAKVASDLYAVAEYYSVDVDRLEKRRITPLANS RYALIERSNMTPYRDIHSALSKITLMTITPVIAIEEDALDENNEKVRRELINGCTY QVSSHVAKSKLFEGERYKFMKRAQYLELDLVHIIASDMHNDSRPHMAEADLVS OKGEAKAOELFDNPKRIWMDLI"
gene	3917..4612
CDS	/gene="cps19a"
gene	3917..4612
CDS	/gene="cps19a"
gene	/function="chain length regulation and export"
CDS	/codon_start=1
gene	/transl_table=11
CDS	/product="Cps19a"
gene	/protein_id="AAC78665.1"
CDS	/db_xref="GI:3907601"
gene	/translation="MMKEONTIEIDVFQILTKMKKLIILVALYTGAGAFSTFI VKPEYSTRIRIYVNRQGDGPGITNODLAGYLVADYEIILSDYLERVATNKL DMPAKLTISKVQVTPADRIIVSISVDKQPEEASRIANSLREVAERKIAIVTSVDY TTLEARPATPSSPNVRNMLVGLGAAVTYTVLLIELDTRVRRPDEVQVLM PLGVPDFDKMK"
gene	4622..5311
CDS	/gene="cps19a"
gene	4622..5311
CDS	/gene="cps19a"
gene	/function="chain length regulation and export"
CDS	/codon_start=1
gene	/transl_table=11
CDS	/product="Cps19a"
gene	/protein_id="AAC78666.1"
CDS	/db_xref="GI:3907602"
gene	/translation="MPTLEISOAKLELVKKAEEYNAICTNLOSGDDLKVFSTTSYK AGEGKTTSTNINAMAFAGYKTLILDADRNSVSGVFSRRTIGLTFELSGTID SOCICDNINENLVIAQGSVSPNPTALDQSNNTNHTETLRKFDYVYDTRAGVYI DAATIMOKCDASILLVTRAGETKRDLQAKAEQLEQTKSCLGAVLNLKDLTSVERKGSY GSYGNRKOK"
gene	5326..6687
CDS	/gene="cps19a"
gene	5326..6687
CDS	/gene="cps19a"
gene	/codon_start=1
gene	/transl_table=11
CDS	/product="glucosyl-1-phosphate transferase Cps19a"
gene	/protein_id="AAC78667.1"
CDS	/db_xref="GI:3907603"
gene	/translation="MSEKLAQPSVAIVOSFLVILLYLLASVRETEIVSTAIVYIL HYFVYISDGRNPFKRGYILEVOTLKILFFALAISSNFFLEDRFSISRGMITF LLHVLVLYLVNPFIRMYMRAYPNRGSKILILTATSVEKVLRLIESDYVGEI VAVSVLDKPDFOHDDLVVAEGEIVNEATHEVDEVEINLPSKYNIGELVSOFTMG IDVTYVNLNAPDMARNOICEMAGLNVVTFSTYKTSYAKRLIDVGLVGLILCG LVSIVLPLIRKDGSAFQOTRIGKGRHFTFYKPSMCMVDAEKKRELMQNTMOG GMFKVDDDPRIITKIGHIRKTSIDELDQFNVLKGMSSLYCTGPPYDEVTEHTPCK RRLSFKPGITGLMQVSGRSEIKNFDEVKLDVAVIDGWTIMKIDIELTKVAVLMD GAK"
gene	6693..7436
CDS	/gene="cps19a"
gene	6693..7436
CDS	/gene="cps19a"
gene	/codon_start=1
gene	/transl_table=11
CDS	/product="putative N-acetyl-mannosamine transferase Cps19a"
gene	/protein_id="AAC78668.1"
CDS	/db_xref="GI:3907604"
gene	/translation="AMERIOIDGITDPELMKFTVDAVEQYVLKKHPLHMGVADKI MOCYEDERKIKIVNESGIIINADGASVLAASKFLGTVPPEVAGIDILMKHLLELSNKG YSVYFPGAKEDVLDQMKIIFERYPALNVGVNGYFSPEDERKHIOEDIKERKDPVF VGIYSPKREYIIOSFMDNGINAVFMGVGGSFVLSGHIKRAPLMOQNAHLEMLFRVAN EPRKLFKRYFVGNVSFLIRKVLRAKRGYKL"
gene	7475..8257
CDS	/gene="cps19a"
gene	7475..8257
CDS	/product="Cps19a"

```

/gene="cps19a"
/function="putative protein involved in formation of
phospho-diester linkage"
/codon_start=1
/transl_table=11
/product="Cps19a"
/protein_id="AAC78669.1"
/db_xref="GI:3907605"
/translation="MDAIKEFOKICKENNDIFELGGSVLAQVKGDFIPMDMDIA
VREGYDKPGIKRIRAGKYOVLAQYCDPLHCYFPLLEDBRRKICPRTNLT
GHLHDIIPDGAHPMSFKRTIFGKVTYRRLASGTYVDHDMHAKKLLTIGF
FKRIEFAKLEPQNVYRRLDNLKRYDMKQYAGTINSLAKEMPEVIEGEEVK
PEDAFKFPTEYDRYKRLXGENTLHEPSDEKSHLGCG"
8259.9137
/gene="cps19a"
8259.9137
/gene="cps19a"
/codon_start=1
/transl_table=11
/product="putative rhamnosyl transferase Cps19a"
/protein_id="AAC78670.1"
/db_xref="GI:3907606"
/translation="MCTYILHLYKYLETSCVKSIRKGNSTKOIIVINDSNNGTG
EKIQELYESDEIDVLIHNNAGFARGNNAVOFAKEXKPPDPMVIMNDIEIEEF
EKIVDIYRKEKEFLGPDIFSTYQLOHNPRLTHYEEVKALNEKKGSOVLA
LKIKCMKSSKVLRTAIONRRKRSKQOVENDILHGSFTYVSRDLKEEYAFN
PNTFVEYELDYAEELKGYKRIYTPRIKYLHONVATNOYTNLVENTLESNKNCF
ESTSYFLMKRNEGV"
9138.10472
/gene="cps19a"
9138.10472
/gene="cps19a"
/codon_start=1
/transl_table=11
/product="polysaccharide polymerase Cps19a"
/protein_id="AAC78671.1"
/db_xref="GI:3907607"
/translation="MTEYELCLTLELFFEFPAFNDLIAPVVMVMEFLISSVFA
LIYVNMNIEXGLAVLILISGIIIVSMPILANSPINTKIKYDRLIDIOEMKIAL
TIYDILITLYRREIRHNALSHGTYGNSFOFPNNATSYEBELVIRISIRLNIID
VSATIFGTFINNETISHKRNKDLILPLIFLISKTLSSGRDITIKILAYVMA
YIQQRKVGMDVISHKYNRLGVLINGIPFYSLFSGSTRVFEESISTYLG
SIQHNQYIONIGVAEFGDESFAIMILGNLGFVYNSVHLEFROLGITMGNY
TEFRPMDHGFGLVAFSEFVYGFALFKLRKSRAGFKDITIIYSYFYWFL
SSIEQYFTTISLFLVFLVLYLMAFFWMLDFHRKLYVILSDTSIKSEN"
10497.11924
/gene="cps19a"
10497.11924
/gene="cps19a"
/codon_start=1
/codon_table=11

Query Match      24.0% Score 31. DB 1: Length 18754;
Best Local Similarity 59.8%; Pred. No. 17;
Matches 52; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

gene
CDS
13888 attactgaacaaagaccgcagcagatgctgtcttctgtgtgctgtga 91
13828 attacagagcttaacaaagccttaacccggtgacaccttctgtgacgtatggg 13887
92 cgtgtgttcacacgcatattatccga 118
13888 cctgtgtttgctgtgtgatacgcga 13914

RESULT 3
AC024405/174118 bp DNA HTG 11-AUG-2001
LOCUS Homo sapiens chromosome 11 clone RP11-574M7 map 11, *** SEQUENCING
IN PROGRESS ***, 1 ordered pieces.
AC024405
AC024405.6 GI:15148297
KEYWORDS HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE human.

```

```

ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS 1 (bases 1 to 174118)
TITLE Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL Homo sapiens chromosome 11, clone RP11-574M7
REFERENCE Unpublished
AUTHORS 2 (bases 1 to 174118)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,
Bouhagalter,B., Brown,A., Burkett,G., Campodiano,A., Castle,A.,
Chapel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M.,
Fenster,J., Ferreira,P., Fitzhugh,M., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Coyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heatord,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Landers,T., Largocque,K., Lehoczy,J., Levine,B.,
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
McEwan,P., McGuirk,A., McKernan,K., McPheeters,R., Meldrum,J.,
Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,T.M.,
Peterson,K., Pierre,N., Pisan,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talmas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
Travers,M., Triggillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
Zody,M.
Direct Submission
TITLE Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome
JOURNAL Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Aug 11, 2001 this sequence version replaced g1:15144553.
All repeats were identified using RepeatMasker.
Smit,A.F.A. & Green,P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIRB
Web site: http://www.seg.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 574_M7
Center clone name: 574_M7
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 174118: contig of 174118 bp in length.
* Location/Qualifiers
source
1. 174118
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11"
/clone="RP11-574M7"
/clone.lib="RP11-574M7 Human Male BAC"
BASE COUNT 54806 a 34652 c 31384 g 53276 t
ORIGIN
Query Match      22.9% Score 29.6; DB 2: Length 174118;
Best Local Similarity 59.5%; Pred. No. 53;
Matches 50; Conservative 0; Mismatches 34; Indels 0; Gaps 0;
25 ttgcagcattactgaacaaagaccgcagcagatgctgtcttctgtgtgctgtga 84
11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11

```

```

Db 76069 TTGGTCACGCTCCTGACAGCTCTTATCCAGATCTGATGCTGCTTGTGCTGCT 76010
OY 85 gctgtgacgtgtgtccacacgta 108
      |||| 11 ||||| 111
Db 76009 CCTGTCCAGTCCTGCTCCAGCTCTTA 75986

RESULT 4
AC069403 185341 bp DNA HTG 22-FEB-2001
LOCUS Homo sapiens chromosome 11 clone RP11-54B5, WORKING DRAFT SEQUENCE,
DEFINITION 2 unordered pieces.
AC069403
AC069403.11 GI:11878497
VERSION AC069403.11
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 185341)
AUTHORS Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
          Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
          Mao,J., Marathe,R., Morehouse,A.J., Oefner,P., Palm,C.J.,
          Oefner,P., Palm,C.J., Ramirez,D., Southwick,A.M., Wilhelmy,J.,
          Yu,S. and Davis,R.W.
          Unpublished
          2 (bases 1 to 185341)
          Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
          Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
          Mao,J., Marathe,R., Morehouse,A.J., Oefner,P., Palm,C.J.,
          Ramirez,D., Wilhelmy,J., Yu,S. and Davis,R.W.
          Direct Submission
          Submitted (29-MAY-2000) DNA Sequencing and Technology Center,
          Stanford University, 855 California Avenue, Palo Alto, CA 94304,
          USA

COMMENT
----- Genome Center
Center: Stanford DNA Sequencing and Technology Development
Center
Center code: SDSTDC
Web site: http://sequence-www.stanford.edu/group/human/
Contact: hum-info@sequence.stanford.edu
----- Project Information
Center project name: RP11-54B5
Center clone name: RP11-54B5
----- Summary Statistics
Sequencing Vector: M13mp18; X02513; 100% of reads
Sequencing Vector: plasmid; plasmid accession: 0% of reads
Chemistry: Dye-terminator; 1% of reads
Chemistry: Dye-terminator Big Dye; 99% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 184714 bases at least Q40
Consensus quality: 184960 bases at least Q30
Consensus quality: 185096 bases at least Q20
Insert size: 113180; agarose-fp
Insert size: 185241; sum-of-contrigs
Quality coverage: 15.2x in Q20 bases; agarose-fp
Quality coverage: 9.3x in Q20 bases; sum-of-contrigs.
NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contrigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contrigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 43623: contig of 43623 bp in length
* 43624 43723: gap of unknown length
* 43724 185341: contig of 141618 bp in length.
Location/Qualifiers
1. 185341
/organism="Homo sapiens"
/db_xref="taxon:9606"

FEATURES
source

Query Match 22.9%; Score 29.6; DB 2; Length 185341;
Best Local Similarity 59.5%; Pred. No. 53;
Matches 50; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

OY 25 ttgagcattactgacacaaagcccgacgagatgctgtttttgtgtgt 84
      |||| 11 ||||| 11111 111111 111111 111111 11111
Db 71459 TTGGTCACGCTCCTGACAGCTCTTATCCAGATCTGATGCTGCTGCT 71518
OY 85 gctgtgacgtgtgtccacacgta 108
      |||| 11 ||||| 111
Db 71519 CCTGTCCAGTCCTGCTCCAGCTCTTA 71542

RESULT 5
AL162413/c
LOCUS AL162413/c
DEFINITION Homo sapiens chromosome 9 clone RP11-109M17, *** SEQUENCING IN
          PROGRESS ***; 11 unordered pieces.
ACCESSION AL162413
VERSION AL162413.14 GI:15131970
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 331109)
AUTHORS Corby,N.
TITLE Direct Submission
JOURNAL Submitted (04-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,
          CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
          requests: clonerequests@sanger.ac.uk
          On Aug 9, 2001 this sequence version replaced gi:15029110.

COMMENT
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: BA109M17
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 326784 bases at least Q40
Consensus quality: 328297 bases at least Q30
Consensus quality: 329246 bases at least Q20
Insert size: 330109; sum-of-contrigs
Insert size: 303557; 19.2% error; agarose-fp
Quality coverage: 6.30x in Q20 bases; sum-of-contrigs quality
coverage: 7.15x in Q20 bases; agarose-fp
NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contrigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contrigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 59586: contig of 59586 bp in length

```

```

* 59587 59686: gap of 100 bp
* 59687 158058: contig of 98372 bp in length
* 158059 158158: gap of 100 bp
* 158159 265019: contig of 106861 bp in length
* 265020 265119: gap of 100 bp
* 265120 282213: contig of 17094 bp in length
* 282214 282313: gap of 100 bp
* 282314 288703: contig of 6390 bp in length
* 288704 288803: gap of 100 bp
* 288804 309775: contig of 20972 bp in length
* 309776 309875: gap of 100 bp
* 309876 312053: contig of 2178 bp in length
* 312054 312153: gap of 100 bp
* 312154 318992: contig of 6839 bp in length
* 318993 319092: gap of 100 bp
* 319093 322077: contig of 2985 bp in length
* 322078 322177: gap of 100 bp
* 322178 328101: contig of 5924 bp in length
* 328102 328201: gap of 100 bp
* 328202 331109: contig of 2908 bp in length.
*
* Location/Qualifiers
*
* 1..331109
*   /organism="Homo sapiens"
*   /db_xref="taxon:9606"
*   /chromosome="9"
*   /clone="RP11-109M17"
*   /clone_1fb="RPC1-11.1"
*   1..59586
*   /note="assembly_fragment:01096"
*   fragment_chain:1
*   /note="assembly_fragment:00965"
*   fragment_chain:1
*   158159..265019
*   /note="assembly_fragment:03121"
*   fragment_chain:2
*   265120..282213
*   /note="assembly_fragment:04814"
*   fragment_chain:2
*   282314..288703
*   /note="assembly_fragment:03747"
*   fragment_chain:3
*   288804..309775
*   /note="assembly_fragment:00694"
*   fragment_chain:3
*   309876..312053
*   /note="assembly_fragment:00673"
*   312154..318992
*   /note="assembly_fragment:01842"
*   319093..322077
*   /note="assembly_fragment:03292"
*   322178..328101
*   /note="assembly_fragment:03601"
*   328202..331109
*   /note="assembly_fragment:04963"
*
BASE COUNT 105642 a 64007 c 60342 g 100117 t 1001 others
ORIGIN

```

```

Query Match 22.9% Score 29.6 DB 2 Length 331109;
Best Local Similarity 59.3% Pred. No. 54;
Matches 50; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

```

```

OY 25 ttccagcattactacagcaaacagccagcagagatgctggggtctttgtgtgt 84
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 174067 TTGCTCCAGTCTCTACCAAGTCTTATCCAGATCTGCATGCTTGTGTCAGTTGTGCTGCT 174008
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 85 gctgtgacgtgtgtgtccacccgta 108
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 174007 CCTGTCCAGTCTCTGTCACAGTCTTA 173984
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 6

```

AC092021/c
LOCUS       AC092021 179759 bp DNA HTG 12-AUG-2001
DEFINITION Homo sapiens chromosome DNK clone RP11-420H12, WORKING DRAFT
SEQUENCE, 7 unordered pieces.
ACCESSION  AC092021
VERSION    AC092021.2 GI:15148366
KEYWORDS   HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
            1 (bases 1 to 179759)
            Waterston,R.H.
REFERENCE  The sequence of Homo sapiens clone
            2 (bases 1 to 179759)
            Waterston,R.H.
REFERENCE  Direct Submission
            Submitted (14-JUN-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
            On Aug 12, 2001 this sequence version replaced g1:14423621.
COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Center project name: H_NH0420H12
----- Summary Statistics -----
Sequencing vector: M13: 5%
Chemistry: Dye-terminator Big Dye; 99% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 165447 bases at least Q40
Consensus quality: 167029 bases at least Q30
Consensus quality: 178591 bases at least Q20
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 10000: contig of 10000 bp in length
* 10001 10100: gap of unknown length
* 10101 11203: contig of 1103 bp in length
* 11204 11303: gap of unknown length
* 11304 21454: contig of 10151 bp in length
* 21455 21554: gap of unknown length
* 21555 44442: contig of 22888 bp in length
* 44443 44542: gap of unknown length
* 44543 83511: contig of 38969 bp in length
* 83512 83611: gap of unknown length
* 83612 178527: contig of 94916 bp in length
* 178528 178627: gap of unknown length
* 178628 179759: contig of 1132 bp in length.
*
* Location/Qualifiers
*
* 1..179759
*   /organism="Homo sapiens"
*   /db_xref="taxon:9606"
*   /chromosome="9"
*   /clone="RP11-420H12"
*   1..10000
*   /note="assembly_name:Contig1"
*   10101..11203
*   /note="assembly_name:Contig14"
*   11304..21454
*   /note="assembly_name:Contig15"

```

```

FEATURES
source
misc_feature
misc_feature
misc_feature
misc_feature
misc_feature

```



```

misc_feature      21555..44442
                  /note="assembly_name:Contig16
                  clone_end:SP6
                  vector_side:left"
misc_feature      44343..83511
                  /note="assembly_name:Contig17
                  clone_end:T7
                  vector_side:left"
misc_feature      83612..178527
                  /note="assembly_name:Contig18"
                  /note="assembly_name:Contig19"
misc_feature      178628..179759
                  /note="assembly_name:Contig19"
BASE COUNT      49208 a 40863 c 39822 g 49254 t 612 others
ORIGIN
Query Match      22.8%; Score 29.4; DB 2; Length 179759;
Best Local Similarity 56.8%; Pred. No. 62;
Matches 54; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
OY 24 ctctgagcattactgacgacaagaccgacgagatgctggggtctttgtgtg 83
Db 30322 CCTCCTGAGTGGCTGACATTTACGACGCCACGACGATGCTGCGCTAATTGTTATTT 30263
OY 84 tgcgtgacgtgtgttccaacgcgtattatccgga 118
Db 30262 TTAGTAGAGATGGGCTTTCACCATGTTAGCCAGGA 30228

RESULT 7
AC091731/c      190919 bp      DNA      HTG      05-JUL-2001
LOCUS      Homo sapiens chromosome UNK clone RP11-343F24, WORKING DRAFT
DEFINITION      AC091731
SEQUENCE, 11 unordered pieces.
AC091731
AC091731.2 GI:14572169
VERSION      HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS      human.
SOURCE      Homo sapiens
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE      1 (bases 1 to 190919)
AUTHORS      Waterston, R.H.
TITLE      The sequence of Homo sapiens clone
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 190919)
AUTHORS      Waterston, R.H.
TITLE      Submitted (25-MAY-2001) Genome Sequencing Center, Washington
JOURNAL      University School of Medicine, 4444 Forest Park Parkway, St. Louis,
              MO 63108, USA
COMMENT      On Jun 27, 2001 this sequence version replaced gi:14196415.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Project Information
Center project name: HNH0343F24

----- Summary Statistics -----
Sequencing vector: M13; &
Chemistry: Dye-Primer ET; % of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 180703 bases at least Q40
Consensus quality: 184821 bases at least Q30
Insert size: 189919; agarose-fp
Quality coverage: 4.39 in Q20 bases; agarose-fp
Quality coverage: 4.73 in Q20 bases; sum-of-coverage

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 18062: contig of 18062 bp in length
* 18063 18162: gap of unknown length
* 18163 41855: contig of 23693 bp in length
* 41856 41955: gap of unknown length
* 41956 74275: contig of 32320 bp in length
* 74276 74375: gap of unknown length
* 74376 150879: contig of 75504 bp in length
* 150880 150979: gap of unknown length
* 150980 152042: contig of 1063 bp in length
* 152043 152142: gap of unknown length
* 152143 154184: contig of 2042 bp in length
* 154185 154284: gap of unknown length
* 154285 157461: contig of 3177 bp in length
* 157462 157561: gap of unknown length
* 157562 162983: contig of 5422 bp in length
* 162984 163083: gap of unknown length
* 163084 166817: contig of 3734 bp in length
* 166818 166917: gap of unknown length
* 166918 171135: contig of 4218 bp in length
* 171136 171235: gap of unknown length
* 171236 190919: contig of 19684 bp in length.
FEATURES
Source
1. 190919
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="UNK"
/clone="RP11-343F24"
1. 18062
/note="assembly_name:Contig10"
18163..41855
/note="assembly_name:Contig11"
41956..74275
/note="assembly_name:Contig12"
74376..150879
/note="assembly_name:Contig13"
150980..152042
/note="assembly_name:Contig13"
152143..154184
/note="assembly_name:Contig13"
154285..157461
/note="assembly_name:Contig14"
157562..162983
/note="assembly_name:Contig15"
163084..166817
/note="assembly_name:Contig16"
166918..171135
/note="assembly_name:Contig17"
171236..190919
/note="assembly_name:Contig18"
190919..42453 c 43073 g 53129 t 1011 others
BASE COUNT 51253 a 42453 c 43073 g 53129 t 1011 others
ORIGIN
Query Match      22.8%; Score 29.4; DB 2; Length 190919;
Best Local Similarity 56.8%; Pred. No. 62;
Matches 54; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
OY 24 ctctgagcattactgacgacaagaccgacgagatgctggggtctttgtgtg 83
Db 142084 CCTCCTGAGTGGCTGACATTTACGACGCCACGACGATGCTGCGCTAATTGTTATTT 142025
OY 84 tgcgtgacgtgtgttccaacgcgtattatccgga 118
Db 142024 TTAGTAGAGATGGGCTTTCACCATGTTAGCCAGGA 141990

```

[illegible]

```

AC025669      Locus
DEFINITION    Mus musculus chromosome 14 clone RP23-5016 map 14, *** SEQUENCING
IN PROGRESS ***; 5 ordered pieces.
AC025669      AC025669
AC025669      HTG: HTGS_PHASE2; HTGS_FUZZTOP; HTGS_ACTIVEFIN.
SOURCE        Mus musculus
ORGANISM      house mouse.
REFERENCE     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS       Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE         1 (bases 1 to 203674)
JOURNAL       Birren,B., Linton,L., Nusbaum,C. and Lander,E.
REFERENCE     Mus musculus chromosome 14, clone RP23-5016
AUTHORS       2 (bases 1 to 203674)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barra,N., Bastien,V., Beda,F.,
Boguslavsky,I., Bouckgeater,B., Brown,A., Burkett,G.,
Campolongo,A., Castle,A., Chappel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearle,A., Chepelov,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferris,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardina,S., Ginde,S., Goeyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Hooton,L.,
Horeland,T.C., Iliev,I., Johnson,K., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoccky,J.,
Levine,R., Lien,C., Liu,G., Locke,K., MacDonald,P., Margus,N.,
McCarthy,M., McEwan,P., McGuck,A., McKernan,K., McPeeters,R.,
Medlir,M., Menus,L., Mihov,T., Miranda,C., Mlangwa,P., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisan,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rotman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessaye,S., Theodore,D., Tirelli,A., Travers,M., Trigilio,J.,
Vassiliev,H., Vriel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,M.J.,
Young,G., Zahoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (12-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 29, 2001 this sequence version replaced g1:15022095.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-----
Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-----
Project Information
Center project name: L8056
Center clone name: 5_O_16
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
* 1 6904: contig of 6904 bp in length
* 6905 7004: gap of 100 bp
* 7005 24083: contig of 17079 bp in length
* 24084 24183: gap of 100 bp
* 24184 74907: contig of 50724 bp in length
* 74908 75007: gap of 100 bp
* 75008 199324: contig of 124317 bp in length
* 199325 199424: gap of 100 bp
* 199425 203674: contig of 4250 bp in length.
*
Location/Qualifiers
1. 203674

```

```

/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="14"
/map="14"
/clone="RP23-5016"
/clone_1lb="RPCT-23"
BASE COUNT      Female Mouse BAC
ORIGIN           59737 a 46593 c 44067 g 52872 t 405 others

```

Query Match	22.88;	Score 29.4;	DB 2;	Length 203674;
Best Local Similarity	58.68;	Pred. No. 62;		
Matches 51; Conservative	0;	Mismatches 36;	Indels 0;	Gaps 0;

Qy	41	gacaaagccccgcagacagatgagccggagctcttttttttggagcgtctgatacaggtgttc	100
Db	161070	GAGGATGAAACGGGGCAGAGCAAGTAGAGAGCTTTCTCTCCTAGACGCTGGGTGGCTTTGTA	161129
Qy	101	caacgcgtattattccgcgactagtctcag	127
Db	161130	CTGGCGTCTATTTAAGGCGTGTAGTAAG	161156

RESULT	10
AL591946	
LOCUS	AL591946 230838 bp DNA HTG 25-JUL-2001
DEFINITION	Mus musculus chromosome 15 clone RP23-290W7, *** SEQUENCING IN
ACCESSION	PROGRAMS ***, in unsorted pieces.
VERSION	AL591946
KEYWORDS	AL591946.9 GI:15022302
SOURCE	HTG; HTGS_PHASE1; HTGS_DRAF; HTGS_FULLTOP.
ORGANISM	house mouse,
	Mus musculus
	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
	1 (bases 1 to 230838)
REFERENCE	Sims, S.
AUTHORS	Direct Submission
TITLE	Submitted (24-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
JOURNAL	CB10 15A. UK. E-mail enquiries: humquerry@sanger.ac.uk Clone
	requests: clonerequests@sanger.ac.uk
COMMENT	On Jul 26, 2001 this sequence version replaced gi:15021038.

```

Web site: http://www.sanger.ac.uk
Contact: humbrey@sanger.ac.uk
----- Project Information -----
Center project name: bmg290M7
----- Summary Statistics -----
Assembly program: XGAP4, version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 227470 bases at least Q40
Consensus quality: 228370 bases at least Q30
Consensus quality: 228935 bases at least Q20
Insert size: 229738; sum-of-contigs
Insert size: 212606; 9.8% error; agarose-fp
Quality coverage: 8.27x in Q20 bases; sum-of-contigs
Quality coverage: 9.20x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```
FEATURES
source      location/Qualifiers
1..230838  /organism="Mus musculus"
           /db_xref="taxon:10090"
           /chromosome="15"
           /clone="RP23-290M7"
           /clone_1b="RPCI-23"
           1..20354
```

```
misc_feature /note="assembly_fragment:04917  
fragment_chain:1"  
20455 .25820  
/note="assembly_fragment:02648  
fragment_chain:1"  
25921 .44695  
/note="assembly_fragment:00262  
fragment_chain:1"  
44796 .78908  
/note="assembly_fragment:05121  
fragment_chain:1"  
79009 .110859  
/note="assembly_fragment:01341  
fragment_chain:1"  
110960 .129541  
/note="assembly_fragment:01437  
fragment_chain:1"  
129642 .147588  
/note="assembly_fragment:03096  
fragment_chain:1"  
147689 .168917  
/note="assembly_fragment:02255  
fragment_chain:1"  
169018 .182699  
/note="assembly_fragment:04888  
fragment_chain:1"  
182800 .188497  
/note="assembly_fragment:00978  
fragment_chain:1"  
188598 .197147  
/note="assembly_fragment:03681  
fragment_chain:1"  
197248 .230838  
/note="assembly_fragment:03785"  
36784 c 62500 g 56499 t 1104 others
```

Query Match	22.8%	Pred. 29.4%	DB 2,	Length 230838;
Best Local Similarity	56.8%	Pred. No. 62;		
Matches	54;	Conservative	0;	Mismatches 41; Indels 0; Gaps 0;
Oy	1	tgaactctgcgcgtagagattactctgacgacattctgacgacaaagaccgcgcgaga	60	
Db 174665		tgaaactctgtgtccacgacgagatgacatcatttatccctttaaaccgacgcacgaga	174724	
Oy	61	tgcgtcgaggtctcttctgtgtgctctgaagtg	95	
Db 174725		tgcctgcgcgtcttgggggtgggggtgagagtcacatg	174759	

RESULT	11
AF176007	
LOCUS	
DEFINITION	AF176007 1877 bp mRNA ROD 19-NOV-1999 Mus musculus domesticus mitochondrial carrier homolog 1 isoform A mRNA; complete cds; nuclear gene for mitochondrial product.
ACCESSION	AF176007
VERSION	AF176007.2 GI:6453808
KEYWORDS	
SOURCE	. western European house mouse. Mus musculis domestics Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;; Mammalia; Eutheria; Rodentia; Sciurognathl.; Muridae; Murinae; Mus
ORGANISM	Jang,J.S., Hahn,Y., Park,C. and Chung,J.H. Identification of an evolutionary conserved mitochondrial carrier family from various organisms Unpublished 2 (bases 1 to 1877) Jang,J.S., Hahn,Y., Park,C. and Chung,J.H. Direct Submission Submitted (06-AUG-1999) Department of Biological Sciences, Korea Advanced Institute of Science and Technology*, 373-I Guseong-dong
JOURNAL	TITLE
REFERENCE	AUTHORS

FEATURES
source

Location/Qualifiers
1.2193
/organism="Rhizobium leguminosarum bv. trifolii"
/strain="TA1"
/db_xref="taxon:386"
41.44
46.11758
/note="ORF1, similar to product encoded by Pseudomonas aeruginosa aprd gene, GenBank Accession Number X64558 and to product encoded by Erwinia chrysanthemi prtd gene, GenBank Accession Number M60395"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAB57756.1"
/db_xref="GI:1172154"
/translation="MGGGLVFTGASALINILYTSFPMLEVDYRVPISPSILAV
LAILALMYCPGAFELRSRLVRVAGALDEMNGRFRALIKAPLKVIGGDLQ
LADPQIRTEPSGMPFAMFDLPMLPYIVICFELPHRPAIGTAIGSLVLAITMTN
OGRTLSKQSDSANMRNAFOTSIRNSEVHAMAGMAEIMDKSEYRTITROA
SDVNGVATLSKIFRIALQSGTLATGALIVOGASSGIIAGSLTSTRALAPVAAI
GNMRFVSAOOSMARLSLKTIPETIPAPLAAPTKQVYEGALSGPAGORLYSD
VSFGIRAGSALGVIGYSASGKSLARAMGIMPTVRSIRLDGALDDMDGALGRHI
GYLDQVLEFSGTVANONICRFAKESPEVYVAAARAVHDLILPLNGEYETEGG
AALSAGORIALALALYGEPLVYLDPEPNSLDEGERALSAIMSVRAGGIYVI
AHRSGVLAFCDFVLMQGRMAIAGPKKEEVLARVSRPEARTPIAERVAOLKVVDDGM
NAAE"
1764.11770
/note="ORF2, similar to product encoded by Pseudomonas aeruginosa aprd gene, GenBank Accession Number X64558 and to product encoded by Erwinia chrysanthemi prtd gene, GenBank Accession Number M60395"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAB57757.1"
/db_xref="GI:1172155"
/translation="VSKVISEKSLNRHNAVGVLSIALVCGIGGMAATTELSAVI
GEGVIVDPKONIOHLTGIVSELVSENHVTAGVLIIRDGTTTRANLSIYESTL
AQLYARARLAEIRIGAESFEVEENTIDLTSTSAOKL"
BASE COUNT 428 a 596 c 684 g 485 t

Query Match 22.6%; Score 29.2; DB 1; Length 2193;
Best Local Similarity 57.8%; Pred. No. 62;
Matches 52; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 28 gagcattctgacgaacacccgacgagatgctggggctctttgtgtgtgct 87
DB 374 GAACCTTCTGCGGTATGGCCGACGACGATCTGCATCTGCCGTGCGGTCT 433
QY 88 gtacagtgtgtccacacgtattatccgg 117
DB 434 ATATGCTATCTCTCTCTCTTCAATCCG 463

RESULT 14
RLRES 5263 bp DNA BCT 29-OCT-1997
LOCUS Rhizobium leguminosarum prsd, prse, ORF3 genes.
DEFINITION X98117
VERSION X98117.1 GI:1806273
KEYWORDS ORF3 gene; prsd gene; prse gene.
SOURCE Rhizobium leguminosarum.
ORGANISM Rhizobium leguminosarum.
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Rhizobiaceae; Rhizobium.
REFERENCE 1 (bases 1 to 5263)
Krol, J. and Skoriupska, A.
Identification of genes in Rhizobium leguminosarum bv. trifolii
whose products are homologues to a family of ATP-binding proteins

JOURNAL Microbiology 143 (Pt 4), 1389-1394 (1997)
MEDLINE 97286554
REFERENCE 2 (bases 1 to 4423)
AUTHORS Krol, J. E.
TITLE Direct Submission
JOURNAL Submitted (24-May-1996) J.E. Krol, Maria Curie Sklodowska
University - Lublin, Dept. of General Microbiology, Akademicka 19,
PL-20 033 Lublin, POLAND
COMMENT On Jan 29, 1997 this sequence version replaced gi:1360123.
FEATURES
source

Location/Qualifiers
1.5263
/organism="Rhizobium leguminosarum"
/sub_species="trifolii"
/strain="TA1"
/db_xref="taxon:384"
/clone_1lb="PARF136"
/clone_2lb="PARF136"
/sub_clone="pue1"
1250.1254
1256.2968
/gene="prsd"
1256.2968
/gene="prsd"
/codon_start=1
/transl_table=11
/protein_id="CAA66799.1"
/db_xref="GI:1360124"
/db_xref="SPTREMBL:057040"
/translation="MGGGLVFTGASALINILYTSFPMLEVDYRVPISPSILAV
LAILALMYCPGAFELRSRLVRVAGALDEMNGRFRALIKAPLKVIGGDLQ
LADPQIRTEPSGMPFAMFDLPMLPYIVICFELPHRPAIGTAIGSLVLAITMTN
OGRTLSKQSDSANMRNAFOTSIRNSEVHAMAGMAEIMDKSEYRTITROA
SDVNGVATLSKIFRIALQSGTLATGALIVOGASSGIIAGSLTSTRALAPVAAI
GNMRFVSAOOSMARLSLKTIPETIPAPLAAPTKQVYEGALSGPAGORLYSD
VSFGIRAGSALGVIGYSASGKSLARAMGIMPTVRSIRLDGALDDMDGALGRHI
GYLDQVLEFSGTVANONICRFAKESPEVYVAAARAVHDLILPLNGEYETEGG
AALSAGORIALALALYGEPLVYLDPEPNSLDEGERALSAIMSVRAGGIYVI
AHRSGVLAFCDFVLMQGRMAIAGPKKEEVLARVSRPEARTPIAERVAOLKVVDDGM
NAAE"
2975.2981
/gene="prse"
2984.4285
2984.4285
/gene="prse"
/codon_start=1
/transl_table=11
/protein_id="CAA66800.1"
/db_xref="GI:1360125"
/db_xref="SPTREMBL:052859"
/translation="MSKVISEKSLNRHNAVGVLSIALVCGIGGMAATTELSAVI
GEGVIVDPKONIOHLTGIVSELVSENHVTAGVLIIRDGTTTRANLSIYESTL
AQLYARARLAEIRIGAESFEVEENTIDLTSTSAOKLDEOKLFTFSRSLIGMS
OLASRKQALAEQIGVAVOINATNDSGLIEEELGIDVLYKKGILVLPFSAGACG
RSPOORPBGNSKSGGEGKALIEIDRSQLEDEDRSEAKULTVEAQLIAYEERAE
RSICIVTSTRADRTSSRAVHTVNGVIDGULMLVPENNELFEAKVATDIDQ
VHVQSDVAFRSAPDQRTTDPVSGEITISADIYKDERTGISYPLRVKPKPESIA
KTIRLPGMPAEVFIKIGDRTVYSILKPLTDOMQHVEFQF"
BASE COUNT 1123 a 1430 c 1553 g 1157 t


```

/standard_name="GC_rich"
/note="Low_Complexity"
intron
5663..10252
/gene="Mtc1"
/number=1
repeat_region
6800..6946
/standard_name="B1_MM"
/note="SINE"
repeat_region
8270..8354
/rpt_family="Alu"
/standard_name="B1_MM"
/note="SINE"
repeat_region
9109..9257
/rpt_family="Alu"
/standard_name="RSINE1"
/note="SINE"
repeat_region
9292..9450
/rpt_family="B4"
/standard_name="B2"
/note="SINE"
repeat_region
9451..9487
/rpt_family="B2"
/rpt_type=tandem
repeat_region
9488..9510
/rpt_unit=taaa
/standard_name="B2"
/note="SINE"
repeat_region
9516..9596
/rpt_family="B2"
/standard_name="PB1D9"
/note="SINE"
repeat_region
9597..9625
/rpt_family="Alu"
/rpt_type=tandem
repeat_region
9639..9804
/rpt_unit=caaaa
/standard_name="RSINE1"
/note="SINE"
exon
10253..10337
/rpt_family="B4"
/gene="Mtc1"
/number=2
intron
10338..12257
/gene="Mtc1"
/number=2
exon
12258..12364
/gene="Mtc1"
/number=3
intron
12365..12657
/gene="Mtc1"
/number=3
exon
12658..12735
/gene="Mtc1"
/number=4
intron
12736..13027
/gene="Mtc1"
/number=4
exon
13028..13085
/gene="Mtc1"
/number=5
intron
13086..13310
/gene="Mtc1"
/number=5
exon
13311..13362
/gene="Mtc1"
/number=6
intron
13363..14350
/gene="Mtc1"
/number=6
exon
14351..14410

```

Query Match 22.6%; Score 29.2; DB 10; length 21984;
 Best Local Similarity 59.8%; Pred. NO. 67;

```

Matches 49; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
QY 48 accccgaccgagatggtcggggtcttltgtgtgtgctgtacgtgtgtccacgt 107
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 16844 ATCCCTCACCTCCTGCGGCATGTGTTTCTGTGTGCGGCTGTAACTGCGCCACTTC 16903
QY 108 attatcggagctagttcacg 129
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 16904 ATCAATGCTTACTGTGCGGACG 16925

```

Search completed: March 28, 2002, 15:20:21
 Job time: 8185 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2002, 15:15:31 ; Search time 240.01 Seconds

(Without alignments)
460.793 Million cell updates/sec

Title: US-09-816-391a-1_COPY_472_600

Perfect score: 129

Sequence: 1 tgaacttctgcgcgtacgca.....tattcgcgactagtcagcg 129

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N.Geneseq.1101.*

- 1: /SIDS2/gcgcdata/geneseq/NA1980.DAT.*
- 2: /SIDS2/gcgcdata/geneseq/NA1981.DAT.*
- 3: /SIDS2/gcgcdata/geneseq/NA1982.DAT.*
- 4: /SIDS2/gcgcdata/geneseq/NA1983.DAT.*
- 5: /SIDS2/gcgcdata/geneseq/NA1984.DAT.*
- 6: /SIDS2/gcgcdata/geneseq/NA1985.DAT.*
- 7: /SIDS2/gcgcdata/geneseq/NA1986.DAT.*
- 8: /SIDS2/gcgcdata/geneseq/NA1987.DAT.*
- 9: /SIDS2/gcgcdata/geneseq/NA1988.DAT.*
- 10: /SIDS2/gcgcdata/geneseq/NA1989.DAT.*
- 11: /SIDS2/gcgcdata/geneseq/NA1990.DAT.*
- 12: /SIDS2/gcgcdata/geneseq/NA1991.DAT.*
- 13: /SIDS2/gcgcdata/geneseq/NA1992.DAT.*
- 14: /SIDS2/gcgcdata/geneseq/NA1993.DAT.*
- 15: /SIDS2/gcgcdata/geneseq/NA1994.DAT.*
- 16: /SIDS2/gcgcdata/geneseq/NA1995.DAT.*
- 17: /SIDS2/gcgcdata/geneseq/NA1996.DAT.*
- 18: /SIDS2/gcgcdata/geneseq/NA1997.DAT.*
- 19: /SIDS2/gcgcdata/geneseq/NA1998.DAT.*
- 20: /SIDS2/gcgcdata/geneseq/NA1999.DAT.*
- 21: /SIDS2/gcgcdata/geneseq/NA2000.DAT.*
- 22: /SIDS2/gcgcdata/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	28.4	22.0	936	AA79329	DNA encoding Arch
2	27.6	21.4	366	AA767819	H. pylori cell env
3	27.6	21.4	366	AA77499	H. pylori inner me
4	27.6	21.4	1422	AA35720	CDNA encoding a pr
5	27.6	21.4	1613	AA35721	CDNA encoding a pr
6	27.6	21.4	2090	AAV6858	Nucleotide sequenc
7	27.6	21.4	2090	AAV38564	Human nuclear prot
8	27.6	21.4	2150	AAH81781	Human differential
9	27.6	21.4	26778	AAH81477	N. meningitidis pa
10	27.6	21.4	349980	AA721612	N. meningitidis pa
11	27.4	21.2	1356	AA521265	Human CDNA sequenc

12	27.4	21.2	1670	AA233333	Human secreted pro
13	27.4	21.2	1878	AAH16020	Human CDNA sequenc
14	27.4	21.2	1951	AA337414	Human secreted pro
15	27.2	21.1	2628	AAH01309	Escherichia coli n
16	26.8	20.8	330	AA444633	Arabidopsis thalia
17	26.8	20.8	364	AA38294	Zea mays DNA fragm
18	26.8	20.8	418	AA46662	Arabidopsis thalia
19	26.4	20.5	927	AA43424	Zea mays DNA fragm
20	26.4	20.5	1364	AA45331	Arabidopsis thalia
21	26.4	20.5	1365	AA39325	Arabidopsis thalia
22	26.2	20.3	446	AA112040	Arabidopsis thalia
23	26.2	20.3	446	AA133374	Probe #1973 for ge
24	26.2	20.3	446	AA101966	Probe #2060 used t
25	26.2	20.3	598	AA140947	Probe #1963 used t
26	26.2	20.3	746	AA292283	Porcine melanocort
27	26.2	20.3	907	AA20670	Human ninturin 2 v
28	26.2	20.3	907	AA20667	Human ninturin 2 e
29	26.2	20.3	927	AA389613	Human secreted pro
30	26.2	20.3	1001	AAV20669	Human ninturin 2 v
31	26	20.2	1173	AA11178	Pseudomonas stutze
32	26	20.2	1192	AA11178	Pseudomonas stutze
33	26	20.2	1228	AA30428	H. pylori inner me
34	26	20.2	1483	AAV30428	H. pylori outer me
35	26	20.2	25801	AAH89885	Human bone marrow
36	26	20.2	43280	AAH1177	Pseudomonas stutze
37	25.8	20.0	1158	AAV84061	Tyrlactone synthase
38	25.8	20.0	1262	AAV06554	Beta-carotene hydr
39	25.6	19.8	1483	AAV63609	Alcaligenes (Deley
40	25.6	19.8	6413	AAV31988	Mouse SMI gene. M
41	25.6	19.8	6604	AAV31981	Human Down syndrom
42	25.4	19.7	413	AA41051	Zea mays DNA fragm
43	25.4	19.7	1068	AAV28385	Human MCP-4 recept
44	25.4	19.7	1426	AAV28385	Human CCR-2 CC che
45	25.4	19.7	1426	AA34773	Human adenosine re

ALIGNMENTS

RESULT 1	AA79329	standard; DNA; 936 BP.
ID	AA79329	standard; DNA; 936 BP.
AC	AA79329	
XX	16-FEB-1998	(first entry)
DE	DNA encoding Archaeoglobus fulgidus esterase VC16-16NC.	
XX	Esterase; thermostable enzyme; ester; chiral compound; cheese;	
KW	pulp; paper; lignin removal; sugar; lignocellulose;	
KW	disease resistance; feedstuff; ss.	
XX	Archaeoglobus fulgidus strain VC16.	
XX	Key	Location/Qualifiers
FT	unsure	313..315
FT	CD5	1..936
FT	/*tag- a	/note= "encodes Ile"
FT	/*tag- b	
FT	/transl_except-	(pos: 382..384, aa:Ala)
FT	/transl_except-	(pos: 385..387, aa:Ala)
FT	/transl_except-	(pos: 430..432, aa:Ala)
FT	/transl_except-	(pos: 487..489, aa:Gly)
FT	/transl_except-	(pos: 499..501, aa:Ala)
FT	/transl_except-	(pos: 505..507, aa:Val)
FT	/transl_except-	(pos: 559..561, aa:Ile)
FT	/transl_except-	(pos: 691..693, aa:Asn)
PN	W09730160-A1.	
XX	21-AUG-1997.	
PD		

[illegible][illegible]

RESULT 5

AAK35721/c

ID AAK35721 standard; CDNA: 1613 BP.

XX AAK35721;

AC 09-JUL-1999 (first entry)

DE cDNA encoding a protein identified by the signal sequence trap method.

XX Signal sequence trap method; SST method; immunisation; inhibition;

XX infection; allergy; cancer; regulation; tissue formation; tissue repair;

XX activin activity; inhibin activity; chemokine activity;

XX cytokine activity; blood coagulation regulation; agonist; antagonist;

XX metabolic disorder; hormonal disorder; immune disorder;

XX severe combined immunodeficiency; SCID; AIDS; thrombosis; cancer;

XX wound; ss.

XX Homo sapiens.

XX MO9918126-A1.

XX 15-APR-1999.

XX 06-OCT-1998; 98WO-JP04514.

XX 07-OCT-1997; 97JP-0274674.

XX (ONKO) ONO PHARM CO LTD.

XX Fukushima D, Shibayama S, Tada H;

XX WPI: 1999-277254/23.

XX P-PSDB: AAY02371.

XX Polypeptides identified by the signal sequence trap method from a

XX human cDNA library

XX Claim 4; Page 141-144; 281pp; Japanese.

XX AAK35694-X35747 represent cDNA sequences that encode novel polypeptides

XX (AAK02358-84) which are identified from a human placental cDNA library

XX by the signal sequence trap (SST) method. The polypeptides have a

XX broad range of physiological activity, including immunisation against

XX and inhibition of infections, allergies and cancer; regulation of tissue

XX formation and repair; activin/inhibin activity; chemokine/cytokine

XX activity; blood coagulation regulation; and receptor/ligand agonist

XX or antagonist activity. The polypeptides can be used for prevention

XX and treatment of disorders including infections by bacteria, yeasts and

XX viruses (including HIV) and protozoa; metabolic and hormonal disorders;

XX immune disorders (including severe combined immunodeficiency (SCID)

XX and AIDS; thrombosis; cancer; and traumatic or surgical wounds.

XX Sequence 1613 BP; 624 A; 243 C; 359 G; 387 T; 0 other:

Query Match 21.4%; Score 27.6; DB 20; Length 1613;

Best Local Similarity 52.6%; Pred. No. 6.4;

Matches 60; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

Oy 7 tctgtcgtagcagattacttcgagcattactgcagacaagaagccgcagcagatgctcg 66

Db 729 TCTGCTGCAATTCATTCTTTTCATGATTTATTCGCTCAATGCTCACTGACTCTCT 670

Oy 67 gggctttttgtgtgctgctgacgtgtgtccaaacgcatattccggact 120

Db 669 GGTGTCAATGATATTCGTGAGTCGAGTGCTCTCTGTTTATTTGTTTCACATT 616

RESULT 6

AAV68588/c

ID AAV68588 standard; DNA: 2090 BP.

XX AAV68588;

XX 16-FEB-1999 (first entry)

DE Nucleotide sequence encoding the human nuclear protein.

XX Human nuclear protein; HEC; cell growth abnormality; hybridisation;

XX amplification; antibody; immunoassay; malignancy; mitosis; antigen;

XX immunoprecipitation; immunisation; vaccine; ss.

XX Homo sapiens.

XX Key

XX Location/Qualifiers

XX CDS 105..2033

XX /tag= a

XX /product= "human nuclear protein"

XX MO9845433-A1.

XX 15-OCT-1998.

XX 03-APR-1998; 98WO-US06727.

XX 04-APR-1997; 97US-0042609.

XX (TEXA) UNIV TEXAS SYSTEM.

XX Alfred DC, Chamness GC, Clark GM, Hilsenbeck SG;

XX Osborn CK;

XX WPI: 1998-594481/50.

XX P-PSDB: AAW80950.

XX New nucleic acid encoding human highly expressed in cancer nuclear

XX protein - used for diagnosis and for modulation of the cell cycle to

XX control malignant and other cell growth abnormalities

XX Claim 1; Pages 55-58; 75pp; English.

XX This is the nucleotide sequence of a human nuclear protein (HEC) used

XX in the method of the invention to diagnose cell growth abnormalities.

XX Fragments of HEC are used to detect HEC-encoding nucleic acid, in

XX usual hybridisation or amplification assays, while antibodies are used

XX in conventional immunoassays to detect HEC or peptides. HEC is used to

XX modulate cell cycle progression (by disrupting chromatid separation)

XX and antibodies are used to disorder sister chromatid alignment and

XX separation in interphase cells, disrupting mitosis, particularly for

XX control of malignancy and other cell growth abnormalities at the

XX mitosis stage. Antibodies can also be used for protein purification, to

XX isolate sequences encoding HEC or related proteins; to study HEC

XX distribution in cells and to isolate antigens by immunoprecipitation.

XX HEC may also be used in vaccines and antibodies for passive

XX immunisation.

XX Sequence 2090 BP; 750 A; 362 C; 464 G; 514 T; 0 other:

Query Match 21.4%; Score 27.6; DB 19; Length 2090;

Best Local Similarity 52.6%; Pred. No. 6.9;

Matches 60; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

Oy 7 tctgtcgtagcagattacttcgagcattactgcagacaagaagccgcagcagatgctcg 66

Db 1239 TCTGCTGCAATTCATTCTTTTCATGATTTATTCGCTCAATGCTCACTGACTCTCT 1180

Oy 67 gggctttttgtgtgctgctgacgtgtgtccaaacgcatattccggact 120

Db 1179 GGTGTCAATGATATTCGTGAGTCGAGTGCTCTCTGTTTATTTGTTTCACATT 1126

RESULT 7

AAV38564/c
ID AAV38564 standard; cDNA: 2090 BP.
AC AAV38564;
XX
XX 26-OCF-1998 (first entry)
XX
XX Human nuclear protein HEC cDNA.
DE
XX HEC gene; highly expressed in cancer; human; nuclear protein;
KW mitosis; cell cycle; cell proliferation; malignancy; vaccine; ss.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FT CDS 105..203
FT /*tag- a
FT polyA_signal 2070..2075
FT /*tag- b
XX
XX MO9827994-A1.
XX
XX PD 02-JUL-1998.
XX
XX PF 18-DEC-1997; 97MO-US23385.
XX
XX PR 20-DEC-1996; 96US-0033600.
XX
XX (TEXA) UNIV TEXAS SYSTEM.
XX
XX PI Chen P, Chen Y, Lee W, Riley DJ;
XX
XX DR MPI: 1998-377401/32.
XX P-PSDB; AAM62596.
XX
XX Nucleic acid encoding human nuclear protein HEC modulating mitosis -
PT useful to, e.g control cell malignancies and other cell growth
PT abnormalities at mitosis stage and to produce HEC protein and
PT peptide(s)
XX
XX PS Claim 2; Page 59-62; 93pp; English.
XX
XX CC This cDNA clone codes for a novel human nuclear protein (see
CC AAM62596) designated HEC (highly expressed in cancer) that appears to
CC be crucial for normal mitosis, possibly in regulation of normal
CC progression of M phase. A 1.8 kb cDNA fragment was originally
CC cloned from a human B cell cDNA library by its interaction with the
CC C-terminus of Rb in a yeast two-hybrid system. This fragment was
CC used as a probe to screen a human B cell cDNA library, and the
CC longest clone obtained included the full-length coding sequence for
CC HEC. HEC nuclear acid segments and their encoded polypeptides are
CC important in modulating mitosis and thus regulating cell
CC proliferation. They are useful in compositions to control cell
CC malignancies and other cell growth abnormalities at the cell
CC mitosis stage. DNA segments encoding HEC are also useful to
CC produce HEC protein/peptides and to detect HEC-encoding sequences
CC (e.g. from related species), by contacting with a nucleic acid
CC sample under suitable conditions and detecting hybridisation
CC (claimed Kits are provided). HEC peptides are useful in vaccines
CC and may be administered to cells to disrupt chromatid separation
CC and so modulate cell cycle progression (claimed).
XX
XX SO Sequence 2090 BP; 750 A; 362 C; 464 G; 514 T; 0 other;

Query Match 21.4%; Score 27.6; DB 19; Length 2090;
Best Local Similarity 52.6%; Pred. No. 6.9;
Matches 60; Conservative 0; Mismatches 54; Indels 0; Gaps 0;
Oy 7 tctgtcgtacgacttctgcagcattactgcagacacacccgcagatgctcg 66
Db 1239 TCTGCTGCAATTCATTCTTTCATGATTTATTCGCTCAATGCTGCACACTGACTTCT 1180

Oy 67 gggtcttttgggtggtgcgtgcagcgtgtgtccaaaccgtattattccggact 120
Db 1179 GGTGTCAATGATATTCGTTCATGATTTATTCGCTCAATGCTGCACACTGACTTCT 1126
RESULT 8
ID AAH81781/c
AC AAH81781;
XX
XX 21-SEP-2001 (first entry)
XX
XX DE Human differential transcription-associated cDNA SEQ ID 290.
XX
XX KW Differential transcription; human; rat; tumour cell; cytosolic;
XX Ras modulator; Class II tumour suppressor gene; gene therapy; ss.
XX
XX OS Homo sapiens.
XX
XX PN MO200157058-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 31-JAN-2001; 2001MO-EP01003.
XX
XX PR 31-JAN-2000; 2000DE-1004102.
XX
XX (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
XX PI Rosenthal A, Hinzmann B, Schaefer R, Zuber J, Tcherntsa O;
XX PI Grips M, Hellriegel M, Schmitz A, Sers C;
XX
XX DR MPI: 2001-483415/52.
XX
XX PT Nucleic acids differentially expressed between tumor and normal cells,
XX PT useful for diagnosis or therapy of tumors and for screening active
XX PT agents
XX
XX PS Disclosure; Page 444-445; 579pp; German.
XX
XX CC This invention describes a nucleic acid (I) with differential expression
XX CC between tumour and normal cells and which has cytostatic activity. (I)
XX CC work as modulators of Ras activity by inducing expression of tumour
XX CC suppressor genes. (I), and polypeptides encoded by them, are useful as
XX CC targets for diagnosis or therapy and in screening to determine the
XX CC effects of an active compound (potential pharmaceutical) on a cell line,
XX CC particularly for diagnosis and treatment of tumors, especially by
XX CC modulating expression of (I) (by gene therapy, antisense RNA or ribozyme
XX CC methods) or by modulating the amount and/or location of (I)-encoded
XX CC polypeptides (by administration of the polypeptide or its activator,
XX CC antibody (optionally as a conjugate) or inhibitor). The method allows
XX CC identification of many Class II tumour suppressor genes (i.e. genes that
XX CC are not primary targets for tumour-initiating mutations).
XX CC AAH81492-AAH83376 represent the human and rat derived nucleic acid
XX CC fragments described in the method of the invention.
XX
XX SO Sequence 2150 BP; 810 A; 362 C; 464 G; 514 T; 0 other;

Query Match 21.4%; Score 27.6; DB 22; Length 2150;
Best Local Similarity 52.6%; Pred. No. 7;
Matches 60; Conservative 0; Mismatches 54; Indels 0; Gaps 0;
Oy 7 tctgtcgtacgacttctgcagcattactgcagacacacccgcagatgctcg 66
Db 1239 TCTGCTGCAATTCATTCTTTCATGATTTATTCGCTCAATGCTGCACACTGACTTCT 1180
Oy 67 gggtcttttgggtggtgcgtgcagcgtgtgtccaaaccgtattattccggact 120
Db 1179 GGTGTCAATGATATTCGTTCATGATTTATTCGCTCAATGCTGCACACTGACTTCT 1126

Db 7636 GACGACCTGCTCGGGCTGATTTTTCGAGGTTTGATCGAAATAAACGATCGAAG 7637
 0y 62 gtcggggctctt 75
 Db 7636 CATCTGTGTGTGT 7623

RESULT 10
 AAF21612
 ID AAF21612 standard; DNA; 349980 BP.
 AC AAF21612;
 DT 13-MAR-2001 (first entry)
 DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:113.
 XX Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
 XX diagnosis; antigen; detection; infection; gene therapy; antibacterial;
 XX ds.
 XX Neisseria meningitidis.
 OS
 PN WO200066791-A1.
 PD 09-NOV-2000.
 PF 08-MAR-2000; 2000MO-US05928.
 PR 30-APR-1999; 99US-0132068.
 PR 08-OCT-1999; 99MO-US23573.
 PR 28-FEB-2000; 2000GB-0004695.
 PA (CHIR) CHIRON CORP.
 PA (GENO) INST GENOMIC RES.
 PI Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Masiarant V,
 PI Galeotti C, Mora M, Ratti G, Scarselli M, Scariato V, Rappuoli R;
 PI Frazer CM, Grandi G;
 DR WPI: 2000-647603/62.
 XX Neisseria meningitidis B full length genome sequence and open reading
 PT frames are used to detect, treat and prevent Neisserial infections -
 XX
 PS
 XX

Claim 7: Appendix A; 692pp; English.

The present invention describes the full length genome of Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607 to AAF21613 represent fragments of the NMB genomic sequence, as the sequence was too long to go in a record on its own it was split into 8 sequences which overlap each other at the beginning and end of each sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to AAF21606 represent PCR primers which are used in the exemplification of the present invention. The NMB genome and fragments from it have antibacterial activity, and can be used in vaccines and gene therapy. Neisseria nucleic acids, proteins and/or antibodies which binds to the proteins can be used in compositions for treating or preventing infection due to Neisserial bacteria or as a diagnostic reagent for detecting the presence of Neisserial bacteria or of antibodies raised to Neisserial bacteria. Computers, computer memory, computer storage medium or computer databases can be used in a search to identify open reading frames (ORFs) or coding sequences within the NMB genome. The DNA sequences provide further opportunities to find antigenic or immunogenic proteins which are more effective in vaccines than the outer membrane proteins currently used.

Sequence 349980 BP; 86473 A; 95646 C; 85908 G; 81953 T; 0 other;
 XX
 SO

PR	31-JUL-1998;	98US-0094935.
PR	10-AUG-1998;	98US-0095880.
PR	11-AUG-1998;	98US-0096068.
PR	06-MAY-1999;	99US-0096068.
XX		
PA	(GENY) GENETICS INST INC.	
PI	Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;	
PI	Merberg D, Treacy M, Agostino MJ, Steinsinger RJ, Bowman MR;	
PI	Diblasio-Smith E, Widom A;	
XX		
DR	WPI; 2000-052937/04.	
DR	P-PSDB; AAY53015.	
XX		
PT	New polynucleotides encoding secreted human proteins, derived from	
PT	adult placenta, adult retina, fetal brain, fetal	
PS	Claim 44; Page 395-396; 492pp; English.	
XX		
XX	The present invention describes new human secreted proteins which were	
CC	isolated from adult placenta, adult retina, foetal brain, foetal kidney,	
CC	adult blood, adult brain, adult thyroid, adult bladder, adult neural	
CC	tissue, adult testes, and adult lymph node cDNA libraries. The human	
CC	secreted proteins, and the polynucleotides encoding them, are predicted	
CC	to have biological activities which would make them suitable for	
CC	treating, preventing or ameliorating medical conditions in humans and	
CC	animals. Suggested activities include nutritional activity, cytokine	
CC	and cell proliferation/differentiation activity, immune stimulating	
CC	(e.g. as vaccines) or suppressing activity, haematopoiesis regulating	
CC	activity, tissue growth activity, activin/inhibin activity,	
CC	chemotactic/chemokinetic activity, hemostatic and thrombolytic	
CC	activity, receptor/ligand activity, anti-inflammatory activity,	
CC	cadherin/tumour invasion suppressor activity, and tumour inhibition	
CC	activity. The polynucleotides are also stated to be useful for gene	
CC	therapy. AA33316 to AA33373 encode human secreted proteins, and	
CC	AA52998 to AA53060 represent human secreted proteins, given in the	
XX	present invention.	
XX		
SQ	Sequence 1670 BP; 362 A; 447 C; 472 G; 389 T; 0 other;	
XX		
Query Match	21.2%; Score 27.4; DB 21; Length 1670;	
Best Local Similarity	54.5%; Pred. No. 7.6;	
Matches	55; Conservative 0; Mismatches 46; Indels 0; Gaps 0	
OY	24 cttcgagcattactcgaagaaagaccgcagacgagatgctggggtcttttggtag 83	
DB	493 ctcgtcggtcttcgttgattatccctccacccctctggcgatgtgttcttctgg 552	
OY	84 tgcgtgacgtgttgcacaacgctattctcgactatt 124	
DB	553 ggcctgaacctcgtggccactcatcattcaatgctactcgtgt 593	
XX		
RESULT 13		
AAH16020		
ID	AAH16020 standard; cDNA; 1878 BP.	
AC	AAH16020;	
XX		
DT	26-JUN-2001 (first entry)	
XX		
DE	Human cDNA sequence SEQ ID NO:14672.	
XX		
XX	Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	EP1074617-A2.	
XX		
PD	07-FEB-2001.	
XX		
PF	28-JUL-2000; 2000EP-0116126.	

XX	29-JUL-1999;	99JP-0248036.	
PR	27-AUG-1999;	99JP-0300253.	
PR	11-JAN-2000;	2000JP-0118776.	
PR	02-MAY-2000;	2000JP-0183767.	
PR	09-JUN-2000;	2000JP-0241899.	
XX			
PA	(HELI-) HELIX RES INST.		
PI	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;		
PI	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;		
XX			
XX	WPI: 2001-318749/34.		
PT	Primer sets for synthesizing polynucleotides, particularly the 5602		
PT	full-length CDNA's defined in the specification, and for the detection		
PT	and/or diagnosis of the abnormality of the proteins encoded by the		
PT	full-length CDNA's -		
PS	Claim 8; SEQ ID 14672; 2537bp + CD ROM; English.		
XX			
XX			
CC	The present invention describes primer sets for synthesizing 5602		
CC	full-length CDNA's defined in the specification. Where a primer set		
CC	comprises: (a) an oligo- α T primer and an oligonucleotide complementary		
CC	to the complementary strand of a polynucleotide which comprises one of		
CC	the 5602 nucleotide sequences defined in the specification, where the		
CC	oligonucleotide comprises at least 15 nucleotides; or (b) a combination		
CC	of an oligonucleotide comprising a sequence complementary to the		
CC	complementary strand of a polynucleotide which comprises a 5'-end		
CC	sequence and an oligonucleotide comprising a sequence complementary to a		
CC	polynucleotide which comprises a 3'-end sequence, where the		
CC	oligonucleotide comprises at least 15 nucleotides and the combination of		
CC	the 5'-end sequence/3'-end sequence is selected from those defined in		
CC	the specification. The primer sets can be used in antisense therapy and		
CC	in gene therapy. The primers are useful for synthesizing polynucleotides		
CC	particularly full-length CDNA's. The primers are also useful for the		
CC	detection and/or diagnosis of the abnormality of the proteins encoded by		
CC	the full-length CDNA's. The primers allow obtaining of the full-length		
CC	CDNA's easily without any specialised methods. AAH03166 to AAH13628 and		
CC	AAH13633 to AAH18742 represent human CDNA sequences; AAB92446 to		
CC	AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632		
CC	represent oligonucleotides, all of which are used in the exemplification		
CC	of the present invention.		
XX			
XX			
SO	Sequence 1878 BP; 356 A; 526 C; 567 G; 429 T; 0 other;		
	Query Match	21.28;	Score 27.4; DB 22; Length 1878;
	Best Local Similarity	54.5%;	Prod. No. 7.8;
	Matches	55;	Conservative 0; Mismatches 46; Indels 0; Gaps
Oy	24 cctgagatcattacgcagacaagaaccgcagcgagctgtcggtctctttgttg 83		
Db	743 cctgcggagatcttcctgctgtgattatcccccactcctctggcgatgtgtttcttg 802		
Oy	84 tgcctgacgctgtgtccaaaccgatatattccggactagt 124		
Db	803 ggctgtaacctgtcgccactcatcatgactactgtg 843		
	RESULT 14		
ID	AAH37414		
XX	AAH37414 standard; cDNA: 1951 BP.		
XX			
AC	AAH37414;		
XX			
DT	06-JUL-1999 (first entry)		
XX			
XX			
DE	Human secreted protein cDNA fragment containing gene 46.		
KW	Human; secreted protein; prevention; treatment; protein therapy;		
KW	gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;		
KW	developmental abnormality; foetal deficiency; blood disorder; lymphoma;		

Query Match 21.18; Score 27.2; DB 22; Length 2628;
 Best Local Similarity 52.7%; Pred. No. 10;
 Matches 59; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 7 tctgctgtaacgattactcgaagcattactgacgacaaagacccgacgagatgctg 66
 |||||
 Db 995 tcaacatggtgcatltgacacatggtcagccgaagatcatgaacctgaagacatcatcg 1054
 |||||

QY 67 gggctcttctgtgtgctgtgacgtgtgtcccaaccgtattattccgga 118
 |||||
 Db 1055 cggcgcttctgtcgaaccgctggaagtggtgaccgcgtactattttcga 1106
 |||||

Search completed: March 28, 2002, 15:16:27
 Job time: 7171 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2002, 15:10:40 ; Search time 112.49 Seconds
(without alignments)
259.718 Million cell updates/sec

Title: US-09-816-391A-1_COPY_472_600
Perfect score: 129
Sequence: 1 tgcattctctgctgtagcga.....tattcgcgactagttacgcg 129

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Issued_patents_MA:*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PTUS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	28.4	22.0	936	2	US-08-602-359A-31	Sequence 31, Appl
2	26.2	20.3	907	3	US-08-672-850-7	Sequence 7, Appl
3	26.2	20.3	907	3	US-08-672-850-11	Sequence 11, Appl
4	26.2	20.3	1001	3	US-08-672-850-10	Sequence 10, Appl
5	26	20.2	43280	2	US-08-804-227C-1	Sequence 1, Appl
6	25.8	20.0	1158	4	US-09-198-092-1	Sequence 1, Appl
7	25.8	20.0	2262	2	US-08-674-887A-5	Sequence 5, Appl
8	25.8	20.0	2262	2	US-08-951-844-5	Sequence 5, Appl
9	25.4	19.7	1586	1	US-08-461-244-1	Sequence 1, Appl
10	25.2	19.5	50341	1	US-08-247-901C-1	Sequence 1, Appl
11	25.2	19.5	50341	2	US-09-075-904-1	Sequence 1, Appl
12	25.2	19.5	52297	4	US-08-426-436-1	Sequence 1, Appl
13	25.2	19.5	52297	4	US-08-705-557-1	Sequence 1, Appl
14	25	19.4	5799	1	US-08-483-101-1	Sequence 1, Appl
15	24.8	19.2	799	1	US-08-458-218-98	Sequence 100, App
16	24.8	19.2	799	2	US-08-450-497-100	Sequence 98, App
17	24.8	19.2	799	2	US-08-486-273A-41	Sequence 100, App
18	24.2	18.8	1883	4	US-08-991-677-3	Sequence 41, Appl
19	23.8	18.4	2326	2	US-08-231-193A-41	Sequence 41, Appl
20	23.8	18.4	2326	2	US-08-486-273A-41	Sequence 41, Appl
21	23.8	18.4	2326	3	US-08-486-273A-41	Sequence 41, Appl
22	23.8	18.4	2326	3	US-08-940-086A-41	Sequence 41, Appl
23	23.8	18.4	3243	2	US-08-231-193A-44	Sequence 44, Appl
24	23.8	18.4	3243	2	US-08-486-273A-44	Sequence 44, Appl
25	23.8	18.4	3243	2	US-08-480-474-44	Sequence 44, Appl
26	23.8	18.4	3243	3	US-08-940-086A-44	Sequence 44, Appl
27	23.8	18.4	3698	2	US-08-231-193A-43	Sequence 43, Appl

28	23.8	18.4	3698	2	US-08-486-273A-43	Sequence 43, Appl
29	23.8	18.4	3698	3	US-08-480-474-43	Sequence 43, Appl
30	23.8	18.4	3698	3	US-08-940-086A-43	Sequence 43, Appl
31	23.8	18.4	4002	2	US-08-231-193A-53	Sequence 53, Appl
32	23.8	18.4	4002	2	US-08-486-273A-53	Sequence 53, Appl
33	23.8	18.4	4002	3	US-08-480-474-53	Sequence 53, Appl
34	23.8	18.4	4002	3	US-08-940-086A-53	Sequence 53, Appl
35	23.8	18.4	4017	2	US-08-231-193A-49	Sequence 49, Appl
36	23.8	18.4	4017	2	US-08-486-273A-49	Sequence 49, Appl
37	23.8	18.4	4017	2	US-08-480-474-49	Sequence 49, Appl
38	23.8	18.4	4017	3	US-08-940-086A-49	Sequence 49, Appl
39	23.8	18.4	4053	2	US-08-231-193A-47	Sequence 47, Appl
40	23.8	18.4	4053	2	US-08-486-273A-47	Sequence 47, Appl
41	23.8	18.4	4053	3	US-08-480-474-47	Sequence 47, Appl
42	23.8	18.4	4053	3	US-08-940-086A-47	Sequence 47, Appl
43	23.8	18.4	4068	2	US-08-231-193A-5	Sequence 5, Appl
44	23.8	18.4	4068	2	US-08-486-273A-5	Sequence 5, Appl
45	23.8	18.4	4068	3	US-08-480-474-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1
US-08-602-359A-31
; Sequence 31, Application US/08602359A
; Patent No. 5942430
GENERAL INFORMATION:
APPLICANT: ROBERTSON, Daniel E.
APPLICANT: MURPHY, Dennis
APPLICANT: REID, John
APPLICANT: MAFELA, Anthony
APPLICANT: LINK, Steven
APPLICANT: SWANSON, Ronald V.
APPLICANT: WARREN, Patrick V.
APPLICANT: KOSMOTKA, Anna
TITLE OF INVENTION: ESTERASES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: FISH & RICHARDSON P.C.
STREET: 4225 EXECUTIVE SQUARE, STE 1400
CITY: LA JOLLA
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/602,359A
FILING DATE: February 16, 1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: HAILE, LISA A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/010001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-678-5070
FAX: 619-678-5099
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 936 NUCLEOTIDES
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: GENOMIC DNA
US-08-602-359A-31

Query Match	22.0%;	Score 28.4;	DB 2;	Length 936;
Best Local Similarity	56.4%;	Pred. No. 0.35;		
Matches 53;	Conservative 0;	Mismatches 41;	Indels 0;	Gaps 0;

```

Oy      2  gacctctcgtctgtagcgattacttcgagcattactgacagaacccgaccgagat 61
          |||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      718 GACCTTGAGAACCTTACCTCCTCGCTGATCATATAACGCCGCAATTACGACCCGCTGAGAGAT 777

```

```
Oy      62  gtcggggtcttttgttgtgtcgtgacgtg 95
          |   |   |||   | |||||   |
Db      78  GAAGGAGAAGTTTCGGGCAGATGCTGAGAAAC  81
```

RESULT 2
US-08-672-

```

1      Sequence 7, Application US/08672850
2      Patent No. 6140117
3      GENERAL INFORMATION:
4      APPLICANT: Milbrandt, Jeffrey
5      APPLICANT: Arai, Toshiyuki
6      TITLE OF INVENTION: NINURIN
7      NUMBER OF SEQUENCES: 34
8      CORRESPONDENCE ADDRESS:
9      ADDRESSEE: Fleh, Hohbach, Test, Albritton & Herbert
10     STREET: Four Embarcadero Center, Suite 3400
11     CITY: San Francisco
12     STATE: California
13     COUNTRY: United States
14     ZIP: 94111
15     COMPUTER READABLE FORM:
16     MEDIUM TYPE: Floppy disk
17     COMPUTER: IBM PC compatible
18     OPERATING SYSTEM: PC-DOS/MS-DOS
19     SOFTWARE: PatentIn Release #1.0, Version #1.30
20     CURRENT APPLICATION DATA:
21     APPLICATION NUMBER: US/08/672,850
22     FILING DATE: 24-JUL-1996
23     CLASSIFICATION: 435
24     ATTORNEY/AGENT INFORMATION:
25     NAME: Silva, Robin M.
26     REGISTRATION NUMBER: 38,304
27     REFERENCE/DOCKET NUMBER: A-65610
28     TELECOMMUNICATION INFORMATION:
29     TELEPHONE: (415) 781-1989
30     TELEFAX: (415) 398-3249
31     INFORMATION FOR SEQ ID NO: 7:
32     SEQUENCE CHARACTERISTICS:
33     LENGTH: 907 base pairs
34     TYPE: nucleic acid
35     STRANDEDNESS: unknown
36     TOPOLOGY: unknown
37     MOLECULE TYPE: DNA (genomic)
38     US-08-672-850-7

```

Query Match	20.3%	Score 26.2;	DB 3;	Length 907;
Best Local Similarity	58.2%	Pred. No. 2.2;		
Matches 46;	Conservative	0;	Mismatches 33;	Indels 0;
			Gaps	0;

QY 11 ctctgtagcattacttctcgagcatctgcagacaagaaccgcgacgatgttcgggt 70
| | | | | | | | | | | | | | | |
Db 434 CCCGAGGCTGTATGAACAACATTGATGCACCACACTGAGAAGAACCAAGATGCTGCTGC 375

```

QY      71 ctttgtgtgtcgt 89
          |||||
Db      374 GTGTGAGCTGGTGAGT 356

```

RESULT

US-08-672-850-11/c
; Sequence 11, Application US/08672850
; Patent No. 6140117

GENERAL INFORMATION:

APPLICANT: Milbrandt, Jeffrey
APPLICANT: Araki, Toshiyuki
TITLE OF INVENTION: NINJURIN
NUMBER OF SEQUENCES: 34

ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco

STATE: California
COUNTRY: United States
ZIP: 94111

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/672,850
FILING DATE: 24-JUL-1996

CLASSIFICATION: 433
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robln M.
REGISTRATION NUMBER: 38,304

REFERENCE/DOCKET NUMBER: A-63610
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249

SEQUENCE CHARACTERISTICS:
LENGTH: 907 base pairs
TYPE: nucleic acid

STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
5-08-672-850-11

Query Match	20.38;	Score 26.2;	DB 3;	Length 907;
Best Local Similarity	58.28;	Pred. No. 2.2;		

11 ctctgtagcgattactctgagcattactgacgacaagaccccgacgagatggtcgggt 70
| | | | | | | | | | | | | | | | | | | | | |
434 ccccaaccctctgaaatgcaaaacattgacgtaccacacgtgaagaagaccagatggcgctgc 375

```

71 cttttgtgtgtgctgt 89
   |||||
374 cttttgtgtgtgctgt 356

```

RESULT 4

3-08-672-850-10/c
Sequence 10, Application US/08672850
Patent No. 6140117

GENERAL INFORMATION:
 APPLICANT: Milbrandt, Jeffere
 APPLICANT: Araki, Toshiyuki
 TITLE OF INVENTION: NINJURIN
 NUMBER OF SEQUENCES: 34

CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
STREET: Four Embarradero Center, Suite 3400
CITY: San Francisco

STATE: California
COUNTRY: United States
ZIP: 94111
COUNTRY: UNITED STATES

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS

OPERATING SYSTEM: PC DOS/MS DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

DATE: 11/11/2019

RESULT 7
US-08-674-887A-5/c

US-08-461-244-1/c
Sequence 1, Application US/08461244
Patent No. 5776729
GENERAL INFORMATION:
APPLICANT: Soppet, Daniel R.
APPLICANT: Yi, Li
APPLICANT: Ruben, Steven M.
APPLICANT: Rosen, Craig A.
TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR HGB632
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: CAPELLA, BYRNE, BAIN, GILFILLAN, CECCHI
ADDRESSEE: STUART & OLSTEIN
STREET: 6 Becker Farm Road
CITY: Roseland
STATE: New Jersey
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patnulin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/461,244
FILING DATE: 05-JUN-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 Inch 1.44 Mb storage diskette
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Word Processor (ASCII)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/075,904
FILING DATE: May 11, 1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/247,901
FILING DATE: May 23, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Bogosian, Elizabeth A
REGISTRATION NUMBER: 39,911
REFERENCE/DOCKET NUMBER: 96700/475

```

      ZIP: 10016
      COMPUTER READABLE FORM:
      MEDIUM TYPE: 3.5 inch 1.44 MB storage diskette
      COMPUTER: IBM PC Compatible
      OPERATING SYSTEM: MS-DOS
      SOFTWARE: WORD PROCESSOR (ASCII)
      CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/09/426,436
      FILING DATE:
      CLASSIFICATION:
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US/08/705,557
      FILING DATE:
      APPLICATION NUMBER: US/08/057,531
      FILING DATE:
      APPLICATION NUMBER: 07/833,431
      FILING DATE: February 7, 1992
      ATTORNEY/AGENT INFORMATION:
      NAME: Pasquajini, Patricia A.
      REGISTRATION NUMBER: 34,894
      REFERENCE/DOCKET NUMBER: 96700/228
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: (212) 697-5995
      TELEFAX: (212) 286-0854 or 286-0082
      TELEX: TWX 710-581-4766
      INFORMATION FOR SEQ ID NO: 1:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 52297
      TYPE: nucleotide
      STRANDEDNESS: single
      TOPOLOGY: linear
      MOLECULE TYPE:
      DESCRIPTION: phage genome sequence
      HYPOTHETICAL: no
      ANTI-SENSE: no
      FRAGMENT TYPE: not applicable.
      ORIGINAL SOURCE:
      ORGANISM: mycobacteriophage L5
      STRAIN: not applicable
      INDIVIDUAL ISOLATE: L5
      DEVELOPMENTAL STAGE: not applicable
      HAPLOTYPE: not applicable
      TISSUE TYPE: not applicable
      CELL TYPE: not applicable
      CELL LINE: not applicable
      ORGANELLE: not applicable
      IMMEDIATE SOURCE: mycobacteriophage L5 particles
      POSITION IN GENOME: entire genome
      FEATURE:
      NAME/KEY:
      LOCATION:
      IDENTIFICATION METHOD:
      OTHER INFORMATION:
      PUBLICATION INFORMATION:
      AUTHORS: Hatfull and Sarkis
      TITLE: DNA Sequence, Structure and Gene
      TITLE: Expression of Mycobacteriophage L5:
      TITLE: A Phage System for Mycobacterial
      TITLE: Genetics
      JOURNAL: Molecular Microbiology
      VOLUME: 7
      PAGES: 395-405
      DATE: 1993
      US-09-426-436-1
  
```



```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

? TITLE: Expression of Mycobacteriophage L5
 ? TITLE: A Phage System for Mycobacterial
 ? TITLE: Genetics
 ?
 ? JOURNAL: Molecular Microbiology
 ? VOLUME: 7
 ? PAGES: 395-405
 ? DATE: 1993
 ?
 ? US-08-705-557-1

Query Match 19.5%; Score 25.2; DB 4; Length 52297;
Best Local Similarity 66.7%; Pred No. 26;
Matches 36; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

54 accgagatgctggggtccttttctgttggtgcctgtgacgtgtgtccaaacct 107
 ||| || |||| | | ||| | |||| ||| | |||
 Db. 1773 ACCGTGACCATCGCGGCACTGCTGATGTGAGCTGGGACGTATTCGCCGCCGT 1826

```

1 RESULT 14
2 US-08-483-101-1
3
4 Sequence 1, Application US/08483101
5 Patent No. 5932715
6
7 GENERAL INFORMATION:
8
9 APPLICANT: Scott, June R.
10
11 APPLICANT: Froehlich, Barbara
12
13 APPLICANT: Caton, Judy
14
15 TITLE OF INVENTION: CS2 Proteins and Coding Sequences
16
17 NUMBER OF SEQUENCES: 17
18
19 CORRESPONDENCE ADDRESS:
20
21 ADDRESSEE: Greenlee and Winner, P.C.
22
23 STREET: 5370 Manhattan Circle, Suite 201
24
25 CITY: Boulder
26
27 STATE: Colorado
28
29 COUNTRY: US
30
31 ZIP: 80303
32
33
34 COMPUTER READABLE FORM:
35
36 MEDIUM TYPE: Floppy disk
37
38 COMPUTER: IBM PC compatible
39
40 OPERATING SYSTEM: PC-DOS/MS-DOS
41
42 SOFTWARE: PatentIn Release #1.0, Version #1.30
43
44 CURRENT APPLICATION DATA:
45
46 APPLICATION NUMBER: US/08/483,101
47
48 FILING DATE: 07-JUN-1995
49
50 CLASSIFICATION: 424
51
52 ATTORNEY/AGENT INFORMATION:
53
54 NAME: Feider, Donna M.
55
56 REGISTRATION NUMBER: 33878
57
58 REFERENCE/DOCKET NUMBER: 6-95
59
60 TELECOMMUNICATION INFORMATION:
61
62 TELEPHONE: (303) 499-8080
63
64 TELEFAX: (303) 499-8089
65
66 INFORMATION FOR SEQ ID NO: 1:
67
68 SEQUENCE CHARACTERISTICS:
69
70 LENGTH: 5798 base pairs
71
72 TYPE: nucleic acid
73
74 STRANDEDNESS: double
75
76 TOPOLOGY: circular
77
78 MOLECULE TYPE: DNA (genomic)
79
80 HYPOTHETICAL: NO
81
82 ANTI-SENSE: NO
83
84 ORIGINAL SOURCE:
85
86 ORGANISM: Escherichia coli
87
88 FEATURE:
89
90 NAME/KEY: CDS
91
92 LOCATION: 499..1215
93
94 FEATURE:
95
96 NAME/KEY: sig_peptide
97
98 LOCATION: 499..552
99
100 FEATURE:
101
102 NAME/KEY: mat_peptide
103
104 LOCATION: 553..1212
105
106 FEATURE:

```


Fri Mar 29 11:21:40 2002

us-09-816-391a-1_copy_472_600.rni

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2002, 14:42:22 ; Search time 2300.01 Seconds

(without alignments)
602.695 Million cell updates/sec

Title: US-09-816-391a-1_COPY_472_600

Perfect score: 129

Sequence: 1 tgacctctgctgctagcga.....tattcgactagtcagcg 129

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST.*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estda:*
7: em_estro:*
8: em_estov:*
9: em_hlc:*
10: gb_est1:*
11: gb_est2:*
12: gb_hlc:*
13: gb_gsa:*
14: em_gsa_fun:*
15: em_gsa_hum:*
16: em_gsa_inv:*
17: em_gsa_pin:*
18: em_gsa_pro:*
19: em_gsa_tod:*
20: em_gsa_vrt:*
21: em_gsa_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	31.8	24.7	600	11	BG817831 ESSU0253
2	30.8	23.9	524	10	AA517086 vH98110.r
3	30.6	23.7	199	10	AJ282728 4A3A-AB-
4	30.6	23.7	677	10	AV745983 AV745983
5	30.2	23.4	546	11	BG817800 ESSU0222
6	30.2	23.4	577	11	BG817800 ESSU0222
7	30.2	23.4	599	11	BG817905 ESSU0327
8	30.2	23.4	657	11	BG817905 ESSU0327
9	30.2	23.4	682	11	BG817854 ESSU0276
10	30.2	23.4	720	11	BG817854 ESSU0276
11	30.2	23.4	1101	13	BG817630 ESSU0052
12	29.6	22.9	187	11	AL108510 Drosophila
					BF841751 RCP-HR107

13	29.6	22.9	903	13	CNS04GSE	AL290039 Tetradon
14	29.4	22.8	485	11	H76502	H76502 18207 Lambd
15	29.4	22.8	1353	11	B1114606	B1114606 602861984
16	29.2	22.6	259	10	A1613789	A1613789 vH98110.y
17	29.2	22.6	345	11	BE945901	BE945901 UI-M-B20
18	29.2	22.6	365	11	W48020	W48020 mc91e03.r1
19	29.2	22.6	406	10	BE651993	BE651993 UI-M-APO
20	29.2	22.6	415	10	AA103400	AA103400 moc24.f02.r
21	29.2	22.6	420	10	BE226847	BE226847 us73d03.y
22	29.2	22.6	534	11	BE914799	BE914799 60167823
23	29.2	22.6	653	11	BG965887	BG965887 602829669
24	29.2	22.6	695	11	B1079667	B1079667 602829669
25	29.2	22.6	763	11	BG922669	BG922669 602829669
26	29.2	22.6	764	11	BF384357	BF384357 602046710
27	29.2	22.6	790	10	AU080797	AU080797 A0080797
28	29.2	22.6	791	11	BF118685	BF118685 601754973
29	29.2	22.6	838	11	BF247036	BF247036 602359684
30	29.2	22.6	848	11	BF532038	BF532038 602073004
31	29.2	22.6	850	11	BF781560	BF781560 602104468
32	29.2	22.6	863	11	BF784104	BF784104 602110059
33	29.2	22.6	908	11	B1248625	B1248625 602993062
34	29.2	22.6	908	11	B1248625	B1248625 602993062
35	29.2	22.5	940	13	A0572855	A0572855 HS-5398.B
36	29.2	22.5	533	11	BG001371	BG001371 OVA-GN014
37	29.2	22.5	1019	13	CNS05L5T	AL342362 Tetradon
38	28.8	22.3	350	13	A2693856	A2693856 ASF-11DB2
39	28.8	22.3	469	10	A1065277	A1065277 TENU2164
40	28.8	22.3	772	10	A1316149	A1316149 UJ25101.Y
41	28.8	22.3	949	10	AL531914	AL531914 AL531914
42	28.6	22.2	886	11	BF339003	BF339003 602036082
43	28.6	22.2	936	10	BE615543	BE615543 601278885
44	28.4	22.0	510	10	AW231767	AW231767 687024C05
45	28.4	22.0	515	10	BE510594	BE510594 946054A06

ALIGNMENTS

RESULT 1

BG817831/c 600 bp mRNA EST 22-MAY-2001
ESSU0253 S.scabiei CDNA library Sarcopites scabiei CDNA clone
SAS0334 5', mRNA sequence.
BG817831
BG817831.1 GI:14188811
EST.
Sarcopites scabiei.
Sarcopites scabiei.
Eukaryotes; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
Acariformes; Sarcopitiformes; Astigmata; Sarcopitidae; Sarcopitidae;

REFERENCE

1 (bases 1 to 600)
Ljunggren,E.L., Nilsson,D., Naslund,K. and Mattsson,J.G.
Expressed sequence tag analysis of the parasitic mite Sarcopites scabiei

JOURNAL

Unpublished (2001)
Contact: Mattsson J.G.
Department of Parasitology (SWEPPAR)
National Veterinary Institute
SE-751 89 Uppsala, Sweden
Tel: +46 18 674120
Fax: +46 18

FEATURES

source

Similar to p1r17405 scavenger receptor or cysteine-rich protein precursor - sea urchin (Strongyloce nitrolos purpuratus).
High quality sequence stop: 600.
Location/Qualifiers
1..600
/organism="Sarcopites scabiei"
/db_xref="taxon:52283"
/clone="SAS0334"
/clone_lib="S.scabiei CDNA library"

Oy	62	g	g	t	c	g	a	g	a	c	t	t	t	t	t	g	t	g	t	g	t	g	c	t	g	a	c	g	t	g	t	c	100
Db	347	G	T	T	A	G	G	C	A	T	C	T	G	G	T	A	G	A	C	T	T	T	T	G	C	A	T	G	C	T	T	C	

FEATURES	source
JOURNAL	Expressed sequence tag analysis of the parasitic mite Sarcopotes scabiei
COMMENT	<p>Unpublished (2001)</p> <p>Contact: Mattsson J.G. Department of Parasitology (SMEPAR) National Veterinary Institute SE-751 89 Uppsala, Sweden Tel: +46 18 674120 Fax: +46 18</p> <p>Email: jens.mattsson@sva.se Similar to IPIRI71405 scavenger recep tor cysteine-rich protein precursor - sea urchin(Strongyloce ntrotus purpuratus). Seq primer: T3 primer High quality sequence stop: 720.</p>
FEATURES	<p>Location/Qualifiers</p> <p>1..720</p>
BASE COUNT	198 a 184 c 163 g 175 t
ORIGIN	<p>XL1-Blue MRF' cells."</p> <p>The primary library was amplified on</p>
Query Match	23.4%; Score 30.2; DB 11; Length 720;
Best Local Similarity	56.6%; Pred. No. 39;
Matches	56; Conservative 0; Mismatches 43; Indels 0; Gaps 0;
OY	2 gaccttcgtccgcgtagagcattacttcgagcattactgacgacaaagaccgccagccagat 61.
Db	590 GACCATTGAGGAGCAGCACCATTGTTGGGGCTGAGACGCCAAGACCCAAATGAGAAAT 531
OY	62 gtcgcggagctcttctgtgtgtgctgctgagcgtgtgttc 100
Db	530 GTTTAGGAGCATCTGTGTGGAGGCTCTTTTGCATGTCTTC 492
RESULT 11	
CNS017UC/c	1101 bp DNA GSS 26-JUL-1999
LOCUS	Drosophila melanogaster genome survey sequence SP6 end of BAC
DEFINITION	BACN37P23 of DrosBAC library from Drosophila melanogaster (fruit
ACCESSION	fly), genomic survey sequence.
VERSION	AL108510
KEYWORDS	AL108510.1 GI:5628814
SOURCE	GSS.
ORGANISM	Fruit fly.
REFERENCE	Plasmod Drosophila melanogaster
AUTHORS	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
TITLE	Pterygota; Neoptera; Endopterygota; Diptera; Brachyera;
JOURNAL	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
COMMENT	<p>1 (bases 1 to 1101)</p> <p>Genoscope.</p> <p>Direct Submission</p> <p>Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :</p> <p>BP 191 91006 EVRY cedex - FRANCE (E-mail : seqret@genoscope.cns.fr</p> <p>- Web : www.genoscope.cns.fr)</p> <p>Determination of this BAC-end sequence was carried out as part of a</p> <p>collaboration with the European Drosophila Genome Project (EDGP) -</p> <p>http://www.edgp.ebi.ac.uk . This Drosophila melanogaster BAC</p> <p>library (Dros BAC) was made by Alain Billand at CEPH (centre</p> <p>d'Etude du Polymorphisme Humain) with funding provided by a MRC</p> <p>project grant. The DNA was prepared from embryos by Alain Bucheton</p> <p>and Genevieve Payan. It has been constructed in the vector</p> <p>peloBAC11.</p>
FEATURES	<p>Location/Qualifiers</p> <p>1..1101</p>


```

/organism="Drosophila melanogaster"
/plasmid="pBelobAC11"
/db_xref="taxon:7227"
/clone_id="DrosBAC"
/clone="BACN37P23"
/note="end : SP6"

```

Query Match	23.48%	Score 30.2;	DB 13;	Length 1101;
Best Local Similarity	53.78%	Pred. No. 41;		
Matches	36;	Conservative	11;	Mismatches 20;
				Indels 0;
				Gaps 0;

Oy 32 attactgacacaagaaccccgacagatggtcgggctctttgtgtggtcgtga 91
 || ||| : : || : ||| || : || |||| : ||||| ||| || : |||
 Db 1100 ATATCTGGSGSMKAWGSGCCCAAAACCAAMGCTTGGGKKYTTTKTGGGT 1041

```
QY      92 cgtgttg 98
          |::|
Db     1040 GGBGKGC 1034
```

RESULT	12
LOCUS	BF841751/c
DEFINITION	BF841751 187 bp mRNA
ACCESSION	RC2-HR1079-051200-015-e02 HR1079 Homo sapiens CDNA, mRNA sequence
VERSION	BF841751
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE	AUTHORS	TITLE	JOURNAL
1 (bases 1 to 187)			
Dias Neto, E., Garcia Correia, R., Verjovski-Almeida, S., Briones, M. R., Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F., Goldman, G. H., Carvalho, A. F., Matsukuma, A., Bala, G. S., Simpson, D. H., Brindley, A. A., de Oliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare, M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and Simpson, A. J.		Shotgun sequencing of the human transcriptome with ORF expressed sequence tags	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

DOUGLAS
FLOC, HALL, ACAD. SCI. U.S.A. 37 (1971) 2431-2430 (2000)
MEDLINE
20202663
Contact: Simpson A.J.G.
COMMENT

COMMENT
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the PABSP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l1-KC2&l2-KC2-HT1079
051200-015-e02&l3=2000-12-05&l4=1)
Seq primer: puc 18 forward
High quality sequence start: 16
High quality sequence stop: 187.
Location/Qualifiers
1. 187

```

1: 1: taxon="Homo sapiens"
2: /db_xref="taxon:9606"
3: /clone_lib="HT1079"
4: /dev_stage="Adult"
5: /note="Organ: head_neck; Vector: puc18; Site.1: Sma1;
6: Site.2: Sma1; A mini-library was made by cloning products
7: derived from ORESTES PCR (0.5. Letters Patent application
8: No. 196,716 - Ludwig Institute for Cancer Research)
9: profiles into the puc 18 vector. Reverse transcription of
10: tissue mRNA and cDNA amplification were performed under
11: low stringency conditions."

```

BASE COUNT	50 a	59 c	45 g	33 t
ORIGIN				

Query Match	22.9%	Score 29.6;	DB 11;	Length 187;
Best Local Similarity	56.0%	Pred. No. 51;		
Matches	56;	Conservative	0;	Mismatches 44;
				Indels 0;
				Gaps 0;

Oy 24 cttcgagcattactgtacgacaagaaccgcgaacgatgctggggtcttlttgttg 83
||| ||| | ||| | ||| ||| ||| ||| |||
Db 144 CTGCTGGAGTCTTCGTGGATTAATCCCTCACCCCTGGGGCATGTGCTTTCTTGTCG 85

Dy 84 tgcctgacagcgttgcaccacgatatattcccgactagt 123
||||| ||||| ||||| ||||| |||||
Db 84 GGCTGTAACTGCTGCGCCACATTCATCATGACCCTAGGT 45

RESULT	13
CNS04GSE	
LOCUS	903 bp DNA
DEFINITION	GSS
	21-MAY-2000
	Tetradon nigroviridis genome survey sequence 77 end of clone
	100201 of library C from Tetradon nigroviridis genomic survey

ACCESSION	AL290039
VERSION	AL290039.1
KEYWORDS	GI:8028619
SOURCE	GSS; genome survey sequence Tetradodon nigroviridis.

ORGANISM	Reference	Year
Tetradodon nigrovittatus	1	1963
Eukarya: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:		
Actinopterygii: Neopterygii: Teleostei: Euteleostei: Neoteleostei:		
Acanthomorpha: Acanthopterygii: Percomorpha: Tetraodontiformes:		
Tetraodontidae: Tetraodon		
1 (phases 1 to 903)		

REFERENCE AUTHORS	TITLE	JOURNAL
1 (bases 1 to 903) Roest-Crolius/H., Jallion/O., Dasilva/C., Fitzmes/C., Fisher/C., Bonneau/L., Billaut/A., Quetier/F., Saurin/W., Bernot/A. and Weissenbach/J.	Characterization and repeat analysis of the compact genome of the freshwater pufferfish <i>Tetraodon nigroviridis</i>	Unpublished 2 (bases 1 to 903) Roest-Crolius/H., Jallion/O., Dasilva/C., Bonneau/L., Fisher/C.

TITLE	JOURNAL	REFERENCE
Human gene number estimate provided by genome wide analysis using Tetradon nigroviridis DNA sequence	Unpublished	3 (bases 1 to 903)

JOURNAL	Unpublished
REFERENCE	3 (bases 1 to 903)
AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (12-APR-2000) to the EMBL/Genbank/DBJ databases
COMMENT	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the tetraodon <i>nitroviridis</i> genome. For more information, please take a look at http://www.genoscope.cns.fr/tetraodon .

FEATURES	source	location/qualifiers
		1..903
		/organism="Tetradodon nigroviridis"
		/db_xref="taxon:99883"
		/clone="109E21"
		/clone_1bp="G"
		/note="Genoscope sequence ID : CBGCI09AC11LP1-end : 177"
BASE COUNT	231 a	195 c 215 g 230 t 32 others
ORIGIN		

Query Match	22.9%	Score	29.6	DB	13	Length	903
Best Local Similarity	53.2%	Pred No.	61				
Matches	50	Conservative	5	Mismatches	39	Indels	0
						Gaps	0

Oy 6 ttcgcgtcgtagcagatctacttctgagcatlctaacyacaagaaccccgacagatggtc 65
 ||| | :| | | | | | | | | | | | | |
Db 462 TTATCCTTTTGTAGTTCCTCCTCCCACTGTAATTGAAAGAACCACGGCTGAGCTGCAC 521

